



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶ : A61K 38/00, 39/00, 39/395, 39/40, C07H 21/04, C07K 1/00, C12N 15/00, C12P 21/06, G01N 33/53	A1	(11) International Publication Number: WO 98/19689 (43) International Publication Date: 14 May 1998 (14.05.98)
(21) International Application Number: PCT/US97/19226 (22) International Filing Date: 27 October 1997 (27.10.97) (30) Priority Data: 60/029,930 1 November 1996 (01.11.96) US (71) Applicants (for all designated States except US): SMITHK-LINE BEECHAM CORPORATION [US/US]; One Franklin Plaza, Philadelphia, PA 19103 (US). SMITHK-LINE BEECHAM PLC [GB/GB]; New Horizons Court, Brentford, Middlesex TW8 9EP (GB). (72) Inventors; and (75) Inventors/Applicants (for US only): BLACK, Michael, Ter-ance [GB/US]; 502 Milhouse Way, Chester Springs, PA 19425 (US). HODGSON, John, Edward [GB/US]; 260 Lapp Road, Malvern, PA 19355 (US). KNOWLES, David, Justin, Charles [GB/GB]; Downsvew House, 45 Cronks Hill Road, Redhill, Surrey RH1 6LY (GB). LONETTO, Michael, Arthur [US/US]; 18 Victoria Circle, Collegeville, PA 19426 (US). NICHOLAS, Richard, Oakley [GB/US]; 355 Carmen Drive, Collegeville, PA 19426 (US). REID, Robert, H., Jr. [US/US]; 8 Pacer Lane, East Norriton, PA 19401 (US).		ZARFOS, Phillip, N. [US/US]; 1907 Yorktown North, Nor-ristown, PA 19403 (US). (74) Agents: GIMMI, Edward, R. et al.; SmithKline Beecham Corporation, Corporate Intellectual Property, UW2220, 709 Swedeland Road, P.O. Box 1539, King of Prussia, PA 19406-0939 (US). (81) Designated States: CA, JP, US, European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE). Published <i>With international search report.</i>
(54) Title: NOVEL CODING SEQUENCES (57) Abstract <p>This invention relates to newly identified Streptococcal polynucleotides, polypeptides encoded by such polynucleotides, the uses of such polynucleotides and polypeptides, as well as the production of such polynucleotides and polypeptides and recombinant host cells transformed with the polynucleotides. This invention also relates to inhibiting the biosynthesis or action of such polynucleotides or polypeptides and to the use of such inhibitors in therapy.</p>		

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

NOVEL CODING SEQUENCES

FIELD OF THE INVENTION

This invention relates to newly identified polynucleotides and polypeptides, and their production and uses, as well as their variants, agonists and antagonists, and their uses. In particular, in these and in other regards, the invention relates to novel polynucleotides and polypeptides set forth in Table 1.

BACKGROUND OF THE INVENTION

The Streptococci make up a medically important genera of microbes known to cause several types of disease in humans, including otitis media, pneumonia and meningitis. Since its isolation more than 100 years ago, *Streptococcus pneumoniae* (herein *S. pneumoniae*) has been one of the more intensively studied microbes. For example, much of our early understanding that DNA is, in fact, the genetic material was predicated on the work of Griffith and of Avery, Macleod and McCarty using this microbe. Despite the vast amount of research with *S. pneumoniae*, many questions concerning the virulence of this microbe remain.

While certain Streptococcal factors associated with pathogenicity have been identified, e.g., capsule polysaccharides, peptidoglycans, pneumolysins, PspA Complement factor H binding component, autolysin, neuraminidase, peptide permeases, hydrogen peroxide, IgA1 protease, the list is certainly not complete. Further very little is known concerning the temporal expression of such genes during infection and disease progression in a mammalian host. Discovering the sets of genes the bacterium is likely to be expressing at the different stages of infection, particularly when an infection is established, provides critical information for the screening and characterization of novel antibacterials which can interrupt pathogenesis. In addition to providing a fuller understanding of known proteins, such an approach will identify previously unrecognised targets.

GUG is used as an initiating nucleotide, rather than ATG, for a significant number of mRNA's in both Gram positive and Gram negative bacteria. Statistics on the frequency of NTG codons in the start codon for several bacterial species are available on line via computer at http://biochem.otago.ac.nz:800/Transterm/home_page.html).

A discussion of initiation codons in *B. subtilis* is set forth in Vellanoweth, RL.1993 in *Bacillus subtilis* and other Gram Positive Bacteria, Biochemistry, Physiology and Molecular Genetics, Sonenshein, Hoch, Losick Eds. Amer. Soc. Microbiol, Washington DC. p. 699-711. Vellenworth indicates a major difference between *B. subtilis* and the gram-negative organisms is in the choice of initiation codon. 91% of the sequenced *E. coli*

genes start with AUG. By contrast, about 30% of *B. subtilis* and other clostridial branch genes start with UUG or GUG. Moreover, CUG functions as a start codon in *B. subtilis*. Mutations of an AUG initiation codon to GUG or UUG often cause decreased expression in *B. subtilis* and *E. coli*. Generally, translation efficiency is higher with AUG initiation codons. A strong Shine-Delgarno ribosome binding site, however, can compensate almost fully for a weak initiation codon. It has been reported that genes with a range of expression levels have initiation codons other than ATG in gram positives (Vellanoeth, RL.1993 in Bacillus subtilis and other Gram Positive Bacteria, Biochemistry, Physiology and Molecular Genetics, Sonenshein, Hoch, Losick Eds. Amer. Soc. Microbiol, Washington DC. p. 699-711).

Provided herein are ORF sequences from genes possessing GUG initiation codons and proteins expressed therefrom and homologues thereto to be used for screening for antimicrobial compounds. Clearly, there is a need for polypeptide and polynucleotide sequences that may be used to screen for antimicrobial compound and which may also be used to determine the roles of such sequences in pathogenesis of infection, dysfunction and disease. There is also need, therefore, for identification and characterization of such sequences which may play a role in preventing, ameliorating or correcting infections, dysfunctions or diseases.

The polypeptides of the invention have amino acid sequence homology to a known protein(s) as set forth in Table 1.

SUMMARY OF THE INVENTION

It is an object of the invention to provide polypeptides that have been identified as novel polypeptides by homology between an amino acid sequence selected from the group consisting of the sequences set out in Table 1 and a known amino acid sequence or sequences of other proteins such as the protein identities listed in Table 1.

It is a further object of the invention to provide polynucleotides that encode novel polypeptides, particularly polynucleotides that encode polypeptides of *Streptococcus pneumoniae*.

In a particularly preferred embodiment of the invention the polynucleotide comprises a region encoding a polypeptide comprising a sequence selected from the group consisting of the sequences set out in Table 1, or a variant of any of these sequences.

In another particularly preferred embodiment of the invention there is a novel protein from *Streptococcus pneumoniae* comprising an amino acid sequence selected from the group consisting of the sequences set out in Table 1, or a variant of any of these sequences.

In accordance with another aspect of the invention there is provided an isolated nucleic acid molecule encoding a mature polypeptide expressible by the *Streptococcus pneumoniae* 0100993 strain contained in the deposited strain.

A further aspect of the invention there are provided isolated nucleic acid molecules encoding a polypeptide of the invention, particularly *Streptococcus pneumoniae* polypeptide, and including mRNAs, cDNAs, genomic DNAs. Further embodiments of the invention include biologically, diagnostically, prophylactically, clinically or therapeutically useful variants thereof, and compositions comprising the same.

In accordance with another aspect of the invention, there is provided the use of a polynucleotide of the invention for therapeutic or prophylactic purposes, in particular genetic immunization. Among the particularly preferred embodiments of the invention are naturally occurring allelic variants of a polypeptide of the invention and polypeptides encoded thereby.

Another aspect of the invention there are provided novel polypeptides of *Streptococcus pneumoniae* as well as biologically, diagnostically, prophylactically, clinically or therapeutically useful variants thereof, and compositions comprising the same.

Among the particularly preferred embodiments of the invention are variants of the polypeptides of the invention encoded by naturally occurring alleles of their genes.

In a preferred embodiment of the invention there are provided methods for producing the aforementioned polypeptides.

In accordance with yet another aspect of the invention, there are provided inhibitors to such polypeptides, useful as antibacterial agents, including, for example, antibodies.

In accordance with certain preferred embodiments of the invention, there are provided products, compositions and methods for assessing expression of the polypeptides and polynucleotides of the invention, treating disease, for example, including, for example, otitis media, conjunctivitis, pneumonia, bacteremia, meningitis, sinusitis, pleural empyema and endocarditis, and most particularly meningitis, such as for example infection of cerebrospinal fluid, assaying genetic variation, and administering a polypeptide or polynucleotide of the invention to an organism to raise an immunological response against a bacteria, especially a *Streptococcus pneumoniae* bacteria.

In accordance with certain preferred embodiments of this and other aspects of the invention there are provided polynucleotides that hybridize to a polynucleotide sequence of the invention, particularly under stringent conditions.

In certain preferred embodiments of the invention there are provided antibodies against polypeptides of the invention.

In other embodiments of the invention there are provided methods for identifying compounds which bind to or otherwise interact with and inhibit or activate an activity of a polypeptide or polynucleotide of the invention comprising: contacting a polypeptide or polynucleotide of the invention with a compound to be screened under conditions to permit binding to or other interaction between the compound and the polypeptide or polynucleotide to assess the binding to or other interaction with the compound, such binding or interaction being associated with a second component capable of providing a detectable signal in response to the binding or interaction of the polypeptide or polynucleotide with the compound; and determining whether the compound binds to or otherwise interacts with and activates or inhibits an activity of the polypeptide or polynucleotide by detecting the presence or absence of a signal generated from the binding or interaction of the compound with the polypeptide or polynucleotide.

In accordance with yet another aspect of the invention, there are provided agonists and antagonists of the polypeptides and polynucleotides of the invention, preferably bacteriostatic or bacteriocidal agonists and antagonists.

In a further aspect of the invention there are provided compositions comprising a polynucleotide or a polypeptide of the invention for administration to a cell or to a multicellular organism.

Various changes and modifications within the spirit and scope of the disclosed invention will become readily apparent to those skilled in the art from reading the following descriptions and from reading the other parts of the present disclosure.

GLOSSARY

The following definitions are provided to facilitate understanding of certain terms used frequently herein.

"Disease(s)" means any bacterial infection, but preferably a streptococcal infection, such as, otitis media, conjunctivitis, pneumonia, bacteremia, meningitis, sinusitis, pleural empyema, endocarditis, meningitis, and infection of cerebrospinal fluid.

"Host cell" is a cell which has been transformed or transfected, or is capable of transformation or transfection by an exogenous polynucleotide sequence.

"Identity," as known in the art, is a relationship between two or more polypeptide sequences or two or more polynucleotide sequences, as determined by comparing the sequences. In the art, "identity" also means the degree of sequence relatedness between polypeptide or polynucleotide sequences, as the case may be, as determined by the match between strings

of such sequences. "Identity" and "similarity" can be readily calculated by known methods, including but not limited to those described in (*Computational Molecular Biology*, Lesk, A.M., ed., Oxford University Press, New York, 1988; *Biocomputing: Informatics and Genome Projects*, Smith, D.W., ed., Academic Press, New York, 1993; *Computer Analysis of Sequence Data*, Part I, Griffin, A.M., and Griffin, H.G., eds., Humana Press, New Jersey, 1994; *Sequence Analysis in Molecular Biology*, von Heinje, G., Academic Press, 1987; and *Sequence Analysis Primer*, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, 1991; and Carillo, H., and Lipman, D., *SIAM J. Applied Math.*, 48: 1073 (1988). Preferred methods to determine identity are designed to give the largest match between the sequences tested. Methods to determine identity and similarity are codified in publicly available computer programs. Preferred computer program methods to determine identity and similarity between two sequences include, but are not limited to, the GCG program package (Devereux, J., et al., *Nucleic Acids Research* 12(1): 387 (1984)), BLASTP, BLASTN, and FASTA (Altschul, S.F. et al., *J. Molec. Biol.* 215: 403-410 (1990)). The BLAST X program is publicly available from NCBI and other sources (*BLAST Manual*, Altschul, S., et al., NCBI NLM NIH Bethesda, MD 20894; Altschul, S., et al., *J. Mol. Biol.* 215: 403-410 (1990)). As an illustration, by a polynucleotide having a nucleotide sequence having at least, for example, 95% "identity" to a reference nucleotide sequence it is intended that the nucleotide sequence of the tested polynucleotide is identical to the reference sequence except that the polynucleotide sequence may include up to five point mutations per each 100 nucleotides of the reference nucleotide sequence. In other words, to obtain a polynucleotide having a nucleotide sequence at least 95% identical to a reference nucleotide sequence, up to 5% of the nucleotides in the reference sequence may be deleted or substituted with another nucleotide, or a number of nucleotides up to 5% of the total nucleotides in the reference sequence may be inserted into the reference sequence. These mutations of the reference sequence may occur at the 5' or 3' terminal positions of the reference nucleotide sequence or anywhere between those terminal positions, interspersed either individually among nucleotides in the reference sequence or in one or more contiguous groups within the reference sequence. Analogously, by a polypeptide having an amino acid sequence having at least, for example, 95% identity to a reference amino acid sequence is intended that the test amino acid sequence of the polypeptide is identical to the reference sequence except that the polypeptide sequence may include up to five amino acid alterations per each 100 amino acids of the reference amino acid. In other words, to obtain a polypeptide having an amino acid sequence at least 95% identical to a reference amino

acid sequence, up to 5% of the amino acid residues in the reference sequence may be — deleted or substituted with another amino acid, or a number of amino acids up to 5% of the total amino acid residues in the reference sequence may be inserted into the reference sequence. These alterations of the reference sequence may occur at the amino or carboxy terminal positions of the reference amino acid sequence or anywhere between those terminal positions, interspersed either individually among residues in the reference sequence or in one or more contiguous groups within the reference sequence.

"Isolated" means altered "by the hand of man" from its natural state, *i.e.*, if it occurs in nature, it has been changed or removed from its original environment, or both. For example, a polynucleotide or a polypeptide naturally present in a living organism is not "isolated," but the same polynucleotide or polypeptide separated from the coexisting materials of its natural state is "isolated", as the term is employed herein.

"Polynucleotide(s)" generally refers to any polyribonucleotide or polydeoxiribonucleotide, which may be unmodified RNA or DNA or modified RNA or DNA. "Polynucleotide(s)" include, without limitation, single- and double-stranded DNA, DNA that is a mixture of single- and double-stranded regions or single-, double- and triple-stranded regions, single- and double-stranded RNA, and RNA that is mixture of single- and double-stranded regions, hybrid molecules comprising DNA and RNA that may be single-stranded or, more typically, double-stranded, or triple-stranded regions, or a mixture of single- and double-stranded regions. In addition, "polynucleotide" as used herein refers to triple-stranded regions comprising RNA or DNA or both RNA and DNA. The strands in such regions may be from the same molecule or from different molecules. The regions may include all of one or more of the molecules, but more typically involve only a region of some of the molecules. One of the molecules of a triple-helical region often is an oligonucleotide. As used herein, the term "polynucleotide(s)" also includes DNAs or RNAs as described above that contain one or more modified bases. Thus, DNAs or RNAs with backbones modified for stability or for other reasons are "polynucleotide(s)" as that term is intended herein. Moreover, DNAs or RNAs comprising unusual bases, such as inosine, or modified bases, such as tritylated bases, to name just two examples, are polynucleotides as the term is used herein. It will be appreciated that a great variety of modifications have been made to DNA and RNA that serve many useful purposes known to those of skill in the art. The term "polynucleotide(s)" as it is employed herein embraces such chemically, enzymatically or metabolically modified forms of polynucleotides, as well as the chemical forms of DNA and RNA characteristic of viruses and cells, including, for

example, simple and complex cells. "Polynucleotide(s)" also embraces short polynucleotides — often referred to as oligonucleotide(s).

"Polypeptide(s)" refers to any peptide or protein comprising two or more amino acids joined to each other by peptide bonds or modified peptide bonds. "Polypeptide(s)" refers to both short chains, commonly referred to as peptides, oligopeptides and oligomers and to longer chains generally referred to as proteins. Polypeptides may contain amino acids other than the 20 gene encoded amino acids. "Polypeptide(s)" include those modified either by natural processes, such as processing and other post-translational modifications, but also by chemical modification techniques. Such modifications are well described in basic texts and in more detailed monographs, as well as in a voluminous research literature, and they are well known to those of skill in the art. It will be appreciated that the same type of modification may be present in the same or varying degree at several sites in a given polypeptide. Also, a given polypeptide may contain many types of modifications. Modifications can occur anywhere in a polypeptide, including the peptide backbone, the amino acid side-chains, and the amino or carboxyl termini. Modifications include, for example, acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphatidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent cross-links, formation of cysteine, formation of pyroglutamate, formylation, gamma-carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, proteolytic processing, phosphorylation, prenylation, racemization, glycosylation, lipid attachment, sulfation, gamma-carboxylation of glutamic acid residues, hydroxylation and ADP-ribosylation, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins, such as arginylation, and ubiquitination. See, for instance, *PROTEINS - STRUCTURE AND MOLECULAR PROPERTIES*, 2nd Ed., T. E. Creighton, W. H. Freeman and Company, New York (1993) and Wold, F., Posttranslational Protein Modifications: Perspectives and Prospects, pgs. 1-12 in *POSTTRANSLATIONAL COVALENT MODIFICATION OF PROTEINS*, B. C. Johnson, Ed., Academic Press, New York (1983); Seifter et al., *Meth. Enzymol.* 182:626-646 (1990) and Rattan et al., *Protein Synthesis: Posttranslational Modifications and Aging*, Ann. N.Y. Acad. Sci. 663: 48-62 (1992). Polypeptides may be branched or cyclic, with or without branching. Cyclic, branched and branched circular polypeptides may result from post-translational natural processes and may be made by entirely synthetic methods, as well.

“Variant(s)” as the term is used herein, is a polynucleotide or polypeptide that differs from a reference polynucleotide or polypeptide respectively, but retains essential properties. A typical variant of a polynucleotide differs in nucleotide sequence from another, reference polynucleotide. Changes in the nucleotide sequence of the variant may or may not alter the amino acid sequence of a polypeptide encoded by the reference polynucleotide. Nucleotide changes may result in amino acid substitutions, additions, deletions, fusions and truncations in the polypeptide encoded by the reference sequence, as discussed below. A typical variant of a polypeptide differs in amino acid sequence from another, reference polypeptide. Generally, differences are limited so that the sequences of the reference polypeptide and the variant are closely similar overall and, in many regions, identical. A variant and reference polypeptide may differ in amino acid sequence by one or more substitutions, additions, deletions in any combination. A substituted or inserted amino acid residue may or may not be one encoded by the genetic code. A variant of a polynucleotide or polypeptide may be a naturally occurring such as an allelic variant, or it may be a variant that is not known to occur naturally. Non-naturally occurring variants of polynucleotides and polypeptides may be made by mutagenesis techniques, by direct synthesis, and by other recombinant methods known to skilled artisans.

DESCRIPTION OF THE INVENTION

Each of polynucleotide and polypeptide sequences provided herein may be used in the discovery and development of antibacterial compounds. Upon expression of the sequences with the appropriate initiation and termination codons the encoded polypeptide can be used as a target for the screening of antimicrobial drugs. Additionally, the DNA sequences encoding preferably the amino terminal regions of the encoded protein or the Shine-Delgarno region can be used to construct antisense sequences to control the expression of the coding sequence of interest. Furthermore, many of the sequences disclosed herein also provide regions upstream and downstream from the encoding sequence. These sequences are useful as a source of regulatory elements for the control of bacterial gene expression. Such sequences are conveniently isolated by restriction enzyme action or synthesized chemically and introduced, for example, into promoter identification strains. These strains contain a reporter structural gene sequence located downstream from a restriction site such that if an active promoter is inserted, the reporter gene will be expressed.

Although each of the sequences may be employed as described above, this invention also provides several means for identifying particularly useful target genes. The

first of these approaches entails searching appropriate databases for sequence matches in related organisms. Thus, if a homologue exists, the Streptococcal-like form of this gene would likely play an analogous role. For example, a Streptococcal protein identified as homologous to a cell surface protein in another organism would be useful as a vaccine candidate. To the extent such homologies have been identified for the sequences disclosed herein they are reported along with the encoding sequence.

Each of the DNA sequences provided herein may be used in the discovery and development of antibacterial compounds. Because each of the sequences contains an open reading frame (ORF) with an appropriate initiation and termination codons, the encoded protein upon expression can be used as a target for the screening of antimicrobial drugs. Additionally, the DNA sequences encoding the amino terminal regions of the encoded protein can be used to construct antisense sequences to control the expression of the coding sequence of interest. Furthermore, many of the sequences disclosed herein also provide regions upstream and downstream from the encoding sequence. These sequences are useful as a source of regulatory elements for the control of bacterial gene expression. Such sequences are conveniently isolated by restriction enzyme action or synthesized chemically and introduced, for example, into promoter identification strains. These strains contain a reporter structural gene sequence located downstream from a restriction site such that if an active promoter is inserted, the reporter gene will be expressed.

It is believed that bacteria possess a number of ways of regulating gene expression levels, especially in subtle degrees, and the interplay between ribosome binding site and initiation codon is utilized for this purpose for these genes. It is also believed that such genes will be important targets for antimicrobial drug discovery, particularly since pathogenesis genes are believed undergo gene expression regulation during in the pathogenesis process. Therefore, the invention provides ORF sequences possessing a GTG (GUG) initiation codon and protein targets expressed therefrom.

Although each of the sequences may be employed as described above, this invention also provides several means for identifying particularly useful target genes. The first of these approaches entails searching appropriate databases for sequence matches in related organisms. Thus, if a homologue exists, the Streptococcal-like form of this gene would likely play an analogous role. For example, a Streptococcal protein identified as homologous to a cell surface protein in another organism would be useful as a vaccine candidate. To the extent such homologies have been identified for the sequences disclosed herein they are reported along with the encoding sequence.

ORF Gene Expression

Recently techniques have become available to evaluate temporal gene expression in bacteria, particularly as it applies to viability under laboratory and infection conditions. A number of methods can be used to identify genes which are essential to survival *per se*, or essential to the establishment/maintenance of an infection. Identification of an ORF unknown by one of these methods yields additional information about its function and permits the selection of such an ORF for further development as a screening target. Briefly, these approaches include:

1) Signature Tagged Mutagenesis (STM): This technique is described by Hensel *et al.*, Science 269: 400-403(1995), the contents of which is incorporated by reference for background purposes. Signature tagged mutagenesis identifies genes necessary for the establishment/maintenance of infection in a given infection model.

The basis of the technique is the random mutagenesis of target organism by various means (e.g., transposons) such that unique DNA sequence tags are inserted in close proximity to the site of mutation. The tags from a mixed population of bacterial mutants and bacteria recovered from an infected hosts are detected by amplification, radiolabeling and hybridisation analysis. Mutants attenuated in virulence are revealed by absence of the tag from the pool of bacteria recovered from infected hosts.

In *Streptococcus pneumoniae*, because the transposon system is less well developed, a more efficient way of creating the tagged mutants is to use the insertion-duplication mutagenesis technique as described by Morrison *et al.*, J. Bacteriol. 159:870 (1984) the contents of which is incorporated by reference for background purposes.

2) In Vivo Expression Technology (IVET): This technique is described by Camilli *et al.*, Proc. Nat'l. Acad. Sci. USA. 91:2634-2638 (1994), the contents of which is incorporated by reference for background purposes. IVET identifies genes up-regulated during infection when compared to laboratory cultivation, implying an important role in infection. ORF identified by this technique are implied to have a significant role in infection establishment/maintenance.

In this technique random chromosomal fragments of target organism are cloned upstream of a promoter-less recombinase gene in a plasmid vector. This construct is introduced into the target organism which carries an antibiotic resistance gene flanked by resolvase sites. Growth in the presence of the antibiotic removes from the population those fragments cloned into the plasmid vector capable of supporting transcription of the recombinase gene and therefore have caused loss of antibiotic resistance. The resistant pool

is introduced into a host and at various times after infection bacteria may be recovered and assessed for the presence of antibiotic resistance. The chromosomal fragment carried by each antibiotic sensitive bacterium should carry a promoter or portion of a gene normally upregulated during infection. Sequencing upstream of the recombinase gene allows identification of the up regulated gene.

3) Differential display: This technique is described by Chuang *et al.*, *J. Bacteriol.* 175:2026-2036 (1993), the contents of which is incorporated by reference for background purposes. This method identifies those genes which are expressed in an organism by identifying mRNA present using randomly-primed RT-PCR. By comparing pre-infection and post infection profiles, genes up and down regulated during infection can be identified and the RT-PCR product sequenced and matched to ORF 'unknowns'.

4) Generation of conditional lethal mutants by transposon mutagenesis: This technique, described by de Lorenzo, V. *et al.*, *Gene* 123:17-24 (1993); Neuwald, A. F. *et al.*, *Gene* 125: 69-73(1993); and Takiff, H. E. *et al.*, *J. Bacteriol.* 174:1544-1553(1992), the contents of which is incorporated by reference for background purposes, identifies genes whose expression are essential for cell viability.

In this technique transposons carrying controllable promoters, which provide transcription outward from the transposon in one or both directions, are generated. Random insertion of these transposons into target organisms and subsequent isolation of insertion mutants in the presence of inducer of promoter activity ensures that insertions which separate promoter from coding region of a gene whose expression is essential for cell viability will be recovered. Subsequent replica plating in the absence of inducer identifies such insertions, since they fail to survive. Sequencing of the flanking regions of the transposon allows identification of site of insertion and identification of the gene disrupted. Close monitoring of the changes in cellular processes/morphology during growth in the absence of inducer yields information on likely function of the gene. Such monitoring could include flow cytometry (cell division, lysis, redox potential, DNA replication), incorporation of radiochemically labeled precursors into DNA, RNA, protein, lipid, peptidoglycan, monitoring reporter enzyme gene fusions which respond to known cellular stresses.

5) Generation of conditional lethal mutants by chemical mutagenesis: This technique is described by Beckwith, J., *Methods in Enzymology* 204: 3-18(1991), the contents of which are incorporated herein by reference for background purposes. In this technique random chemical mutagenesis of target organism, growth at

temperature other than physiological temperature (permissive temperature) and subsequent replica plating and growth at different temperature (e.g. 42°C to identify ts, 25°C to identify cs) are used to identify those isolates which now fail to grow (conditional mutants). As above close monitoring of the changes upon growth at the non-permissive temperature yields information on the function of the mutated gene. Complementation of conditional lethal mutation by library from target organism and sequencing of complementing gene allows matching with unknown ORF.

6) RT-PCR: *Streptococcus pneumoniae* messenger RNA is isolated from bacterial infected tissue e.g. 48 hour murine lung infections, and the amount of each mRNA species assessed by reverse transcription of the RNA sample primed with random hexanucleotides followed by PCR with gene specific primer pairs. The determination of the presence and amount of a particular mRNA species by quantification of the resultant PCR product provides information on the bacterial genes which are transcribed in the infected tissue. Analysis of gene transcription can be carried out at different times of infection to gain a detailed knowledge of gene regulation in bacterial pathogenesis allowing for a clearer understanding of which gene products represent targets for screens for novel antibacterials. Because of the gene specific nature of the PCR primers employed it should be understood that the bacterial mRNA preparation need not be free of mammalian RNA. This allows the investigator to carry out a simple and quick RNA preparation from infected tissue to obtain bacterial mRNA species which are very short lived in the bacterium (in the order of 2 minute halflives). Optimally the bacterial mRNA is prepared from infected murine lung tissue by mechanical disruption in the presence of TRIzol (GIBCO-BRL) for very short periods of time, subsequent processing according to the manufacturers of TRIzol reagent and DNAase treatment to remove contaminating DNA. Preferably the process is optimised by finding those conditions which give a maximum amount of *Streptococcus pneumoniae* 16S ribosomal RNA as detected by probing Northern blots with a suitably labelled sequence specific oligonucleotide probe. Typically a 5' dye labelled primer is used in each PCR primer pair in a PCR reaction which is terminated optimally between 8 and 25 cycles. The PCR products are separated on 6% polyacrylamide gels with detection and quantification using GeneScanner (manufactured by ABI).

Each of these techniques may have advantages or disadvantage depending on the particular application. The skilled artisan would choose the approach that is the most relevant with the particular end use in mind.

Use of the of these technologies when applied to the ORFs of the present invention enables identification of bacterial proteins expressed during infection, inhibitors of which would have utility in anti-bacterial therapy.

The invention relates to novel polypeptides and polynucleotides as described in greater detail below. In particular, the invention relates to polypeptides and polynucleotides of *Streptococcus pneumoniae*, which is related by amino acid sequence homology to known polypeptide as set forth in Table 1. The invention relates especially to compounds having the nucleotide and amino acid sequence selected from the group consisting of the sequences set out in Table 1, and to the nucleotide sequences of the DNA in the deposited strain and amino acid sequences encoded thereby.

Deposited materials

The deposit has been made under the terms of the Budapest Treaty on the International Recognition of the Deposit of Micro-organisms for Purposes of Patent Procedure. The strain will be irrevocably and without restriction or condition released to the public upon the issuance of a patent. The deposit is provided merely as convenience to those of skill in the art and is not an admission that a deposit is required for enablement, such as that required under 35 U.S.C. §112.

A deposit containing a *Streptococcus pneumoniae* bacterial strain has been deposited with the National Collections of Industrial and Marine Bacteria Ltd. (NCIMB), 23 St. Machar Drive, Aberdeen AB2 1RY, Scotland on 11 April 1996 and assigned NCIMB Deposit No. 40794. The *Streptococcus pneumoniae* bacterial strain deposit is referred to herein as "the deposited bacterial strain" or as "the DNA of the deposited bacterial strain."

The deposited material is a bacterial strain that contains the full length FabH DNA, referred to as "NCIMB 40794" upon deposit.

The sequence of the polynucleotides contained in the deposited material, as well as the amino acid sequence of the polypeptide encoded thereby, are controlling in the event of any conflict with any description of sequences herein.

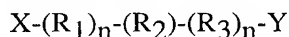
A license may be required to make, use or sell the deposited materials, and no such license is hereby granted.

The deposited strain contains the full length genes comprising the polynucleotides set forth in Table 1. The sequence of the polynucleotides contained in the deposited strain, as well as the amino acid sequence of the polypeptide encoded thereby, are controlling in the event of any conflict with any description of sequences herein.

Polypeptides

The polypeptides of the invention include the polypeptides set forth in Table 1 (in particular the mature polypeptide) as well as polypeptides and fragments, particularly those which have the biological activity of a polypeptide of the invention, and also those which have at least 50%, 60% or 70% identity to a polypeptide sequence selected from the group consisting of the sequences set out in Table 1 or the relevant portion, preferably at least 80% identity to a polypeptide sequence selected from the group consisting of the sequences set out in Table 1, and more preferably at least 90% similarity (more preferably at least 90% identity) to a polypeptide sequence selected from the group consisting of the sequences set out in Table 1, and still more preferably at least 95% similarity (still more preferably at least 95% identity) to a polypeptide sequence selected from the group consisting of the sequences set out in Table 1, and also include portions of such polypeptides with such portion of the polypeptide generally containing at least 30 amino acids and more preferably at least 50 amino acids.

The invention also includes polypeptides of the formula:



wherein, at the amino terminus, X is hydrogen, and at the carboxyl terminus, Y is hydrogen or a metal, R_1 and R_3 are any amino acid residue, n is an integer between 1 and 2000, and R_2 is an amino acid sequence of the invention, particularly an amino acid sequence selected from the group set forth in Table 1. In the formula above R_2 is oriented so that its amino terminal residue is at the left, bound to R_1 , and its carboxy terminal residue is at the right, bound to R_3 . Any stretch of amino acid residues denoted by either R group, where R is greater than 1, may be either a heteropolymer or a homopolymer, preferably a heteropolymer. In preferred embodiments n is an integer between 1 and 1000 or 2000.

A fragment is a variant polypeptide having an amino acid sequence that entirely is the same as part but not all of the amino acid sequence of the aforementioned polypeptides. As with polypeptides, fragments may be "free-standing," or comprised within a larger polypeptide of which they form a part or region, most preferably as a single continuous region, a single larger polypeptide.

Preferred fragments include, for example, truncation polypeptides having a portion of the amino acid sequence of Table 1, or of variants thereof, such as a continuous series of residues that includes the amino terminus, or a continuous series of residues that includes the carboxyl terminus. Degradation forms of the polypeptides of the invention in a host cell, particularly a *Streptococcus pneumoniae*, are also preferred. Further preferred are fragments characterized by structural or functional attributes such as fragments that comprise alpha-helix and alpha-helix forming regions, beta-sheet and beta-sheet-forming regions, turn and turn-forming regions, coil

and coil-forming regions, hydrophilic regions, hydrophobic regions, alpha amphipathic regions, — beta amphipathic regions, flexible regions, surface-forming regions, substrate binding region, and high antigenic index regions.

Also preferred are biologically active fragments which are those fragments that mediate activities of polypeptides of the invention, including those with a similar activity or an improved activity, or with a decreased undesirable activity. Also included are those fragments that are antigenic or immunogenic in an animal, especially in a human. Particularly preferred are fragments comprising receptors or domains of enzymes that confer a function essential for viability of *Streptococcus pneumoniae* or the ability to initiate, or maintain cause disease in an individual, particularly a human.

Variants that are fragments of the polypeptides of the invention may be employed for producing the corresponding full-length polypeptide by peptide synthesis; therefore, these variants may be employed as intermediates for producing the full-length polypeptides of the invention.

In addition to the standard single and triple letter representations for amino acids, the term "X" or "Xaa" is also used. "X" and "Xaa" mean that any of the twenty naturally occurring amino acids may appear at such a designated position in the polypeptide sequence.

Polynucleotides

The nucleotide sequences disclosed herein can be obtained by synthetic chemical techniques known in the art or can be obtained from *S. pneumoniae* 0100993 by probing a DNA preparation with probes constructed from the particular sequences disclosed herein. Alternatively, oligonucleotides derived from a disclosed sequence can act as PCR primers in a process of PCR-based cloning of the sequence from a bacterial genomic source. It is recognised that such sequences will also have utility in diagnosis of the stage of infection and type of infection the pathogen has attained.

To obtain the polynucleotide encoding the protein using the DNA sequence given herein typically a library of clones of chromosomal DNA of *S.pneumoniae* 0100993 in *E. coli* or some other suitable host is probed with a radiolabelled oligonucleotide, preferably a 17mer or longer, derived from the partial sequence. Clones carrying DNA identical to that of the probe can then be distinguished using high stringency washes. By sequencing the individual clones thus identified with sequencing primers designed from the original sequence it is then possible to extend the sequence in both directions to determine the full gene sequence. Conveniently such sequencing is performed using denatured double stranded DNA prepared from a plasmid clone. Suitable techniques are described by

Maniatis, T., Fritsch, E.F. and Sambrook, J. in MOLECULAR CLONING, A Laboratory Manual, 2nd edition, 1989, Cold Spring Harbor Laboratory (see: Screening By Hybridization 1.90 and Sequencing Denatured Double-Stranded DNA Templates 13.70).

Moreover, another aspect of the invention relates to isolated polynucleotides that encode the polypeptides of the invention having a deduced amino acid sequence selected from the group consisting of the sequences in Table 1 and polynucleotides closely related thereto and variants thereof.

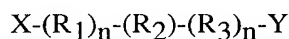
Using the information provided herein, such as the polynucleotide sequences set out in Table 1, a polynucleotide of the invention encoding polypeptide may be obtained using standard cloning and screening methods, such as those for cloning and sequencing chromosomal DNA fragments from bacteria using *Streptococcus pneumoniae* 0100993 cells as starting material, followed by obtaining a full length clone. For example, to obtain a polynucleotide sequence of the invention, such as a sequence set forth in Table 1, typically a library of clones of chromosomal DNA of *Streptococcus pneumoniae* 0100993 in *E.coli* or some other suitable host is probed with a radiolabeled oligonucleotide, preferably a 17-mer or longer, derived from a partial sequence. Clones carrying DNA identical to that of the probe can then be distinguished using stringent conditions. By sequencing the individual clones thus identified with sequencing primers designed from the original sequence it is then possible to extend the sequence in both directions to determine the full gene sequence. Conveniently, such sequencing is performed using denatured double stranded DNA prepared from a plasmid clone. Suitable techniques are described by Maniatis, T., Fritsch, E.F. and Sambrook et al., *MOLECULAR CLONING, A LABORATORY MANUAL*, 2nd Ed.; Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York (1989). (see in particular Screening By Hybridization 1.90 and Sequencing Denatured Double-Stranded DNA Templates 13.70). Illustrative of the invention, the polynucleotides set out in Table 1 were discovered in a DNA library derived from *Streptococcus pneumoniae* 0100993.

The DNA sequences set out in Table 1 each contains at least one open reading frame encoding a protein having at least about the number of amino acid residues set forth in Table 1. The start and stop codons of each open reading frame (herein "ORF") DNA are the first three and the last three nucleotides of each polynucleotide set forth in Table 1.

Certain polynucleotides and polypeptides of the invention are structurally related to known proteins as set forth in Table 1. These proteins exhibit greatest homology to the homologue listed in Table 1 from among the known proteins.

The invention provides a polynucleotide sequence identical over its entire length to each coding sequence in Table 1. Also provided by the invention is the coding sequence for the mature polypeptide or a fragment thereof, by itself as well as the coding sequence for the mature polypeptide or a fragment in reading frame with other coding sequence, such as those encoding a leader or secretory sequence, a pre-, or pro- or prepro- protein sequence. The polynucleotide may also contain non-coding sequences, including for example, but not limited to non-coding 5' and 3' sequences, such as the transcribed, non-translated sequences, termination signals, ribosome binding sites, sequences that stabilize mRNA, introns, polyadenylation signals, and additional coding sequence which encode additional amino acids. For example, a marker sequence that facilitates purification of the fused polypeptide can be encoded. In certain embodiments of the invention, the marker sequence is a hexa-histidine peptide, as provided in the pQE vector (Qiagen, Inc.) and described in Gentz *et al.*, *Proc. Natl. Acad. Sci., USA* 86: 821-824 (1989), or an HA tag (Wilson *et al.*, *Cell* 37: 767 (1984). Polynucleotides of the invention also include, but are not limited to, polynucleotides comprising a structural gene and its naturally associated sequences that control gene expression.

The invention also includes polynucleotides of the formula:



wherein, at the 5' end of the molecule, X is hydrogen, and at the 3' end of the molecule, Y is hydrogen or a metal, R_1 and R_3 is any nucleic acid residue, n is an integer between 1 and 3000, and R_2 is a nucleic acid sequence of the invention, particularly a nucleic acid sequence selected from the group set forth in Table 1. In the polynucleotide formula above R_2 is oriented so that its 5' end residue is at the left, bound to R_1 , and its 3' end residue is at the right, bound to R_3 . Any stretch of nucleic acid residues denoted by either R group, where R is greater than 1, may be either a heteropolymer or a homopolymer, preferably a heteropolymer. In a preferred embodiment n is an integer between 1 and 1000, or 2000 or 3000.

The term "polynucleotide encoding a polypeptide" as used herein encompasses polynucleotides that include a sequence encoding a polypeptide of the invention, particularly a bacterial polypeptide and more particularly a polypeptide of the *Streptococcus pneumoniae* having an amino acid sequence set out in Table 1. The term also encompasses polynucleotides that include a single continuous region or discontinuous regions encoding the polypeptide (for example, interrupted by integrated phage or an insertion sequence or editing) together with additional regions, that also may contain coding and/or non-coding sequences.

The invention further relates to variants of the polynucleotides described herein that encode for variants of the polypeptide having the deduced amino acid sequence of Table 1.

Variants that are fragments of the polynucleotides of the invention may be used to synthesize full-length polynucleotides of the invention.

Further particularly preferred embodiments are polynucleotides encoding polypeptide variants, that have the amino acid sequence of a polypeptide of Table 1 in which several, a few, 5 to 10, 1 to 5, 1 to 3, 2, 1 or no amino acid residues are substituted, deleted or added, in any combination. Especially preferred among these are silent substitutions, additions and deletions, that do not alter the properties and activities of such polynucleotide.

Further preferred embodiments of the invention are polynucleotides that are at least 50%, 60% or 70% identical over their entire length to a polynucleotide encoding a polypeptide having the amino acid sequence set out in Table 1, and polynucleotides that are complementary to such polynucleotides. Alternatively, most highly preferred are polynucleotides that comprise a region that is at least 80% identical over its entire length to a polynucleotide encoding a polypeptide of the deposited strain and polynucleotides complementary thereto. In this regard, polynucleotides at least 90% identical over their entire length to the same are particularly preferred, and among these particularly preferred polynucleotides, those with at least 95% are especially preferred. Furthermore, those with at least 97% are highly preferred among those with at least 95%, and among these those with at least 98% and at least 99% are particularly highly preferred, with at least 99% being the more preferred.

A preferred embodiment is an isolated polynucleotide comprising a polynucleotide sequence selected from the group consisting of: a polynucleotide having at least a 50% identity to a polynucleotide encoding a polypeptide comprising the amino acid sequence of Table 1 and obtained from a prokaryotic species other than *S. pneumoniae*; and a polynucleotide encoding a polypeptide comprising an amino acid sequence which is at least 50% identical to the amino acid sequence of Table 1 and obtained from a prokaryotic species other than *S. pneumoniae*.

Preferred embodiments are polynucleotides that encode polypeptides that retain substantially the same biological function or activity as the mature polypeptide encoded by the DNA of Table 1.

The invention further relates to polynucleotides that hybridize to the herein above-described sequences. In this regard, the invention especially relates to polynucleotides that hybridize under stringent conditions to the herein above-described polynucleotides. As herein used, the terms "stringent conditions" and "stringent hybridization conditions" mean hybridization will occur only if there is at least 95% and preferably at least 97% identity between the sequences. An example of stringent hybridization conditions is overnight incubation at 42°C in a solution comprising: 50% formamide, 5x SSC (150mM NaCl, 15mM trisodium

citrate), 50 mM sodium phosphate (pH7.6), 5x Denhardt's solution, 10% dextran sulfate, and 20 micrograms/ml denatured, sheared salmon sperm DNA, followed by washing the hybridization support in 0.1x SSC at about 65°C. Hybridization and wash conditions are well known and exemplified in Sambrook, *et al.*, Molecular Cloning: A Laboratory Manual, Second Edition, Cold Spring Harbor, N.Y., (1989), particularly Chapter 11 therein.

The invention also provides a polynucleotide consisting essentially of a polynucleotide sequence obtainable by screening an appropriate library containing the complete gene for a polynucleotide sequence set forth in Table 1 under stringent hybridization conditions with a probe having the sequence of said polynucleotide sequence or a fragment thereof; and isolating said DNA sequence. Fragments useful for obtaining such a polynucleotide include, for example, probes and primers described elsewhere herein.

As discussed additionally herein regarding polynucleotide assays of the invention, for instance, polynucleotides of the invention as discussed above, may be used as a hybridization probe for RNA, cDNA and genomic DNA to isolate full-length cDNAs and genomic clones encoding a polypeptide and to isolate cDNA and genomic clones of other genes that have a high sequence similarity to a polynucleotide set forth in Table 1. Such probes generally will comprise at least 15 bases. Preferably, such probes will have at least 30 bases and may have at least 50 bases. Particularly preferred probes will have at least 30 bases and will have 50 bases or less.

For example, the coding region of each gene that comprises or is comprised by a polynucleotide set forth in Table 1 may be isolated by screening using a DNA sequence provided in Table 1 to synthesize an oligonucleotide probe. A labeled oligonucleotide having a sequence complementary to that of a gene of the invention is then used to screen a library of cDNA, genomic DNA or mRNA to determine which members of the library the probe hybridizes to.

The polynucleotides and polypeptides of the invention may be employed, for example, as research reagents and materials for discovery of treatments of and diagnostics for disease, particularly human disease, as further discussed herein relating to polynucleotide assays.

Polynucleotides of the invention that are oligonucleotides derived from the a polynucleotide or polypeptide sequence set forth in Table 1 may be used in the processes herein as described, but preferably for PCR, to determine whether or not the polynucleotides identified herein in whole or in part are transcribed in bacteria in infected tissue. It is recognized that such sequences will also have utility in diagnosis of the stage of infection and type of infection the pathogen has attained.

The invention also provides polynucleotides that may encode a polypeptide that is the mature protein plus additional amino or carboxyl-terminal amino acids, or amino acids interior to

the mature polypeptide (when the mature form has more than one polypeptide chain, for instance). Such sequences may play a role in processing of a protein from precursor to a mature form, may allow protein transport, may lengthen or shorten protein half-life or may facilitate manipulation of a protein for assay or production, among other things. As generally is the case *in vivo*, the additional amino acids may be processed away from the mature protein by cellular enzymes.

A precursor protein, having the mature form of the polypeptide fused to one or more prosequences may be an inactive form of the polypeptide. When prosequences are removed such inactive precursors generally are activated. Some or all of the prosequences may be removed before activation. Generally, such precursors are called proproteins.

In addition to the standard A, G, C, T/U representations for nucleic acid bases, the term "N" is also used. "N" means that any of the four DNA or RNA bases may appear at such a designated position in the DNA or RNA sequence, except it is preferred that N is not a base that when taken in combination with adjacent nucleotide positions, when read in the correct reading frame, would have the effect of generating a premature termination codon in such reading frame.

In sum, a polynucleotide of the invention may encode a mature protein, a mature protein plus a leader sequence (which may be referred to as a preprotein), a precursor of a mature protein having one or more prosequences that are not the leader sequences of a preprotein, or a preproprotein, which is a precursor to a proprotein, having a leader sequence and one or more prosequences, which generally are removed during processing steps that produce active and mature forms of the polypeptide.

Vectors, host cells, expression

The invention also relates to vectors that comprise a polynucleotide or polynucleotides of the invention, host cells that are genetically engineered with vectors of the invention and the production of polypeptides of the invention by recombinant techniques. Cell-free translation systems can also be employed to produce such proteins using RNAs derived from the DNA constructs of the invention.

For recombinant production, host cells can be genetically engineered to incorporate expression systems or portions thereof or polynucleotides of the invention. Introduction of a polynucleotide into the host cell can be effected by methods described in many standard laboratory manuals, such as Davis et al., *BASIC METHODS IN MOLECULAR BIOLOGY*, (1986) and Sambrook et al., *MOLECULAR CLONING: A LABORATORY MANUAL*, 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1989), such as, calcium

phosphate transfection, DEAE-dextran mediated transfection, transvection, microinjection, cationic lipid-mediated transfection, electroporation, transduction, scrape loading, ballistic introduction and infection.

Representative examples of appropriate hosts include bacterial cells, such as streptococci, staphylococci, enterococci *E. coli*, streptomyces and *Bacillus subtilis* cells; fungal cells, such as yeast cells and *Aspergillus* cells; insect cells such as *Drosophila* S2 and *Spodoptera* Sf9 cells; animal cells such as CHO, COS, HeLa, C127, 3T3, BHK, 293 and Bowes melanoma cells; and plant cells.

A great variety of expression systems can be used to produce the polypeptides of the invention. Such vectors include, among others, chromosomal, episomal and virus-derived vectors, *e.g.*, vectors derived from bacterial plasmids, from bacteriophage, from transposons, from yeast episomes, from insertion elements, from yeast chromosomal elements, from viruses such as baculoviruses, papova viruses, such as SV40, vaccinia viruses, adenoviruses, fowl pox viruses, pseudorabies viruses and retroviruses, and vectors derived from combinations thereof, such as those derived from plasmid and bacteriophage genetic elements, such as cosmids and phagemids. The expression system constructs may contain control regions that regulate as well as engender expression. Generally, any system or vector suitable to maintain, propagate or express polynucleotides and/or to express a polypeptide in a host may be used for expression in this regard. The appropriate DNA sequence may be inserted into the expression system by any of a variety of well-known and routine techniques, such as, for example, those set forth in Sambrook *et al.*, *MOLECULAR CLONING, A LABORATORY MANUAL*, (*supra*).

For secretion of the translated protein into the lumen of the endoplasmic reticulum, into the periplasmic space or into the extracellular environment, appropriate secretion signals may be incorporated into the expressed polypeptide. These signals may be endogenous to the polypeptide or they may be heterologous signals.

Polypeptides of the invention can be recovered and purified from recombinant cell cultures by well-known methods including ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography, and lectin chromatography. Most preferably, high performance liquid chromatography is employed for purification. Well known techniques for refolding protein may be employed to regenerate active conformation when the polypeptide is denatured during isolation and or purification.

Diagnostic Assays

This invention is also related to the use of the polynucleotides of the invention for use as diagnostic reagents. Detection of such polynucleotides in a eukaryote, particularly a mammal, and especially a human, will provide a diagnostic method for diagnosis of a disease. Eukaryotes (herein also "individual(s)"), particularly mammals, and especially humans, infected with an organism comprising a gene of the invention may be detected at the nucleic acid level by a variety of techniques.

Nucleic acids for diagnosis may be obtained from an infected individual's cells and tissues, such as bone, blood, muscle, cartilage, and skin. Genomic DNA may be used directly for detection or may be amplified enzymatically by using PCR or other amplification technique prior to analysis. RNA or cDNA may also be used in the same ways. Using amplification, characterization of the species and strain of prokaryote present in an individual, may be made by an analysis of the genotype of the prokaryote gene. Deletions and insertions can be detected by a change in size of the amplified product in comparison to the genotype of a reference sequence. Point mutations can be identified by hybridizing amplified DNA to labeled polynucleotide sequences of the invention. Perfectly matched sequences can be distinguished from mismatched duplexes by RNase digestion or by differences in melting temperatures. DNA sequence differences may also be detected by alterations in the electrophoretic mobility of the DNA fragments in gels, with or without denaturing agents, or by direct DNA sequencing. See, *e.g.*, Myers et al., *Science*, 230: 1242 (1985). Sequence changes at specific locations also may be revealed by nuclease protection assays, such as RNase and S1 protection or a chemical cleavage method. See, *e.g.*, Cotton et al., *Proc. Natl. Acad. Sci., USA*, 85: 4397-4401 (1985).

Cells carrying mutations or polymorphisms in the gene of the invention may also be detected at the DNA level by a variety of techniques, to allow for serotyping, for example. For example, RT-PCR can be used to detect mutations. It is particularly preferred to use RT-PCR in conjunction with automated detection systems, such as, for example, GeneScan. RNA or cDNA may also be used for the same purpose, PCR or RT-PCR. As an example, PCR primers complementary to a nucleic acid encoding a polypeptide of the invention can be used to identify and analyze mutations. These primers may be used for, among other things, amplifying a DNA of the invention isolated from a sample derived from an individual. The primers may be used to amplify the gene isolated from an infected individual such that the gene may then be subject to various techniques for elucidation of the DNA sequence. In this way, mutations in the DNA sequence may be detected and used to diagnose infection and to serotype and/or classify the infectious agent.

The invention further provides a process for diagnosing disease, preferably bacterial infections, more preferably infections by *Streptococcus pneumoniae*, and most preferably disease, comprising determining from a sample derived from an individual a increased level of expression of polynucleotide having the sequence of Table 1. Increased or decreased expression of a polynucleotide of the invention can be measured using any one of the methods well known in the art for the quantitation of polynucleotides, such as, for example, amplification, PCR, RT-PCR, RNase protection, Northern blotting and other hybridization methods.

In addition, a diagnostic assay in accordance with the invention for detecting over-expression of a polypeptide of the invention compared to normal control tissue samples may be used to detect the presence of an infection, for example. Assay techniques that can be used to determine levels of a protein, in a sample derived from a host are well-known to those of skill in the art. Such assay methods include radioimmunoassays, competitive-binding assays, Western Blot analysis and ELISA assays.

Antibodies

The polypeptides of the invention or variants thereof, or cells expressing them can be used as an immunogen to produce antibodies immunospecific for such polypeptides. "Antibodies" as used herein includes monoclonal and polyclonal antibodies, chimeric, single chain, simianized antibodies and humanized antibodies, as well as Fab fragments, including the products of an Fab immunoglobulin expression library.

Antibodies generated against the polypeptides of the invention can be obtained by administering the polypeptides or epitope-bearing fragments, analogues or cells to an animal, preferably a nonhuman, using routine protocols. For preparation of monoclonal antibodies, any technique known in the art that provides antibodies produced by continuous cell line cultures can be used. Examples include various techniques, such as those in Kohler, G. and Milstein, C., *Nature* 256: 495-497 (1975); Kozbor *et al.*, *Immunology Today* 4: 72 (1983); Cole *et al.*, pg. 77-96 in *MONOCLONAL ANTIBODIES AND CANCER THERAPY*, Alan R. Liss, Inc. (1985).

Techniques for the production of single chain antibodies (U.S. Patent No. 4,946,778) can be adapted to produce single chain antibodies to polypeptides of this invention. Also, transgenic mice, or other organisms such as other mammals, may be used to express humanized antibodies.

Alternatively phage display technology may be utilized to select antibody genes with binding activities towards the polypeptide either from repertoires of PCR amplified v-genes of lymphocytes from humans screened for possessing recognition of a polypeptide of

the invention or from naive libraries (McCafferty, J. et al., (1990), *Nature* **348**, 552-554; Marks, J. et al., (1992) *Biotechnology* **10**, 779-783). The affinity of these antibodies can also be improved by chain shuffling (Clackson, T. et al., (1991) *Nature* **352**, 624-628).

If two antigen binding domains are present each domain may be directed against a different epitope - termed 'bispecific' antibodies.

The above-described antibodies may be employed to isolate or to identify clones expressing the polypeptides to purify the polypeptides by affinity chromatography.

Thus, among others, antibodies against a polypeptide of the invention may be employed to treat disease.

Polypeptide variants include antigenically, epitopically or immunologically equivalent variants that form a particular aspect of this invention. The term "antigenically equivalent derivative" as used herein encompasses a polypeptide or its equivalent which will be specifically recognized by certain antibodies which, when raised to the protein or polypeptide according to the invention, interfere with the immediate physical interaction between pathogen and mammalian host. The term "immunologically equivalent derivative" as used herein encompasses a peptide or its equivalent which when used in a suitable formulation to raise antibodies in a vertebrate, the antibodies act to interfere with the immediate physical interaction between pathogen and mammalian host.

The polypeptide, such as an antigenically or immunologically equivalent derivative or a fusion protein thereof is used as an antigen to immunize a mouse or other animal such as a rat or chicken. The fusion protein may provide stability to the polypeptide. The antigen may be associated, for example by conjugation, with an immunogenic carrier protein for example bovine serum albumin (BSA) or keyhole limpet haemocyanin (KLH). Alternatively a multiple antigenic peptide comprising multiple copies of the protein or polypeptide, or an antigenically or immunologically equivalent polypeptide thereof may be sufficiently antigenic to improve immunogenicity so as to obviate the use of a carrier.

Preferably, the antibody or variant thereof is modified to make it less immunogenic in the individual. For example, if the individual is human the antibody may most preferably be "humanized"; where the complementarity determining region(s) of the hybridoma-derived antibody has been transplanted into a human monoclonal antibody, for example as described in Jones, P. et al. (1986), *Nature* **321**, 522-525 or Tempest et al., (1991) *Biotechnology* **9**, 266-273.

The use of a polynucleotide of the invention in genetic immunization will preferably employ a suitable delivery method such as direct injection of plasmid DNA into

muscles (Wolff et al., Hum Mol Genet 1992, 1:363, Manthorpe et al., Hum. Gene Ther. 1993:4, 419), delivery of DNA complexed with specific protein carriers (Wu et al., J Biol Chem. 1989: 264,16985), coprecipitation of DNA with calcium phosphate (Benvenisty & Reshef, PNAS, 1986:83,9551), encapsulation of DNA in various forms of liposomes (Kaneda et al., Science 1989:243,375), particle bombardment (Tang et al., Nature 1992, 356:152, Eisenbraun et al., DNA Cell Biol 1993, 12:791) and *in vivo* infection using cloned retroviral vectors (Seeger et al., PNAS 1984:81,5849).

Antagonists and agonists - assays and molecules

Polypeptides of the invention may also be used to assess the binding of small molecule substrates and ligands in, for example, cells, cell-free preparations, chemical libraries, and natural product mixtures. These substrates and ligands may be natural substrates and ligands or may be structural or functional mimetics. See, *e.g.*, Coligan *et al.*, *Current Protocols in Immunology* 1(2): Chapter 5 (1991).

The invention also provides a method of screening compounds to identify those which enhance (agonist) or block (antagonist) the action of a polypeptides or polynucleotides of the invention, particularly those compounds that are bacteriostatic and/or bacteriocidal. The method of screening may involve high-throughput techniques. For example, to screen for agonists or antagonists, a synthetic reaction mix, a cellular compartment, such as a membrane, cell envelope or cell wall, or a preparation of any thereof, comprising a polypeptide of the invention and a labeled substrate or ligand of such polypeptide is incubated in the absence or the presence of a candidate molecule that may be an agonist or antagonist of a polypeptide of the invention. The ability of the candidate molecule to agonize or antagonize a polypeptide of the invention is reflected in decreased binding of the labeled ligand or decreased production of product from such substrate. Molecules that bind gratuitously, *i.e.*, without inducing the effects of a polypeptide of the invention are most likely to be good antagonists. Molecules that bind well and increase the rate of product production from substrate are agonists. Detection of the rate or level of production of product from substrate may be enhanced by using a reporter system. Reporter systems that may be useful in this regard include but are not limited to colorimetric labeled substrate converted into product, a reporter gene that is responsive to changes in polynucleotide or polypeptide activity, and binding assays known in the art.

Another example of an assay for antagonists of polypeptides of the invention is a competitive assay that combines any such polypeptide and a potential antagonist with a compound which binds such polypeptide, natural substrates or ligands, or substrate or ligand mimetics, under appropriate conditions for a competitive inhibition assay. A polypeptide of the

invention can be labeled, such as by radioactivity or a colorimetric compound, such that the number of such polypeptide molecules bound to a binding molecule or converted to product can be determined accurately to assess the effectiveness of the potential antagonist.

Potential antagonists include small organic molecules, peptides, polypeptides and antibodies that bind to a polynucleotide or polypeptide of the invention and thereby inhibit or extinguish its activity. Potential antagonists also may be small organic molecules, a peptide, a polypeptide such as a closely related protein or antibody that binds the same sites on a binding molecule, such as a binding molecule, without inducing activities induced by a polypeptide of the invention, thereby preventing the action of such polypeptide by excluding it from binding.

Potential antagonists include a small molecule that binds to and occupies the binding site of the polypeptide thereby preventing binding to cellular binding molecules, such that normal biological activity is prevented. Examples of small molecules include but are not limited to small organic molecules, peptides or peptide-like molecules. Other potential antagonists include antisense molecules (see Okano, *J. Neurochem.* 56: 560 (1991); *OLIGODEOXYNUCLEOTIDES AS ANTISENSE INHIBITORS OF GENE EXPRESSION*, CRC Press, Boca Raton, FL (1988), for a description of these molecules). Preferred potential antagonists include compounds related to and variants of a polypeptide of the invention.

Each of the DNA sequences provided herein may be used in the discovery and development of antibacterial compounds. The encoded protein, upon expression, can be used as a target for the screening of antibacterial drugs. Additionally, the DNA sequences encoding the amino terminal regions of the encoded protein or Shine-Delgarno or other translation facilitating sequences of the respective mRNA can be used to construct antisense sequences to control the expression of the coding sequence of interest.

The invention also provides the use of the polypeptide, polynucleotide or inhibitor of the invention to interfere with the initial physical interaction between a pathogen and mammalian host responsible for sequelae of infection. In particular the molecules of the invention may be used: in the prevention of adhesion of bacteria, in particular gram positive bacteria, to mammalian extracellular matrix proteins on in-dwelling devices or to extracellular matrix proteins in wounds; to block protein-mediated mammalian cell invasion by, for example, initiating phosphorylation of mammalian tyrosine kinases (Rosenshine *et al.*, *Infect. Immun.* 60:2211 (1992); to block bacterial adhesion between mammalian extracellular matrix proteins and bacterial proteins that mediate tissue damage and; to block the normal progression of pathogenesis in infections initiated other than by the implantation of in-dwelling devices or by other surgical techniques.

The antagonists and agonists of the invention may be employed, for instance, to inhibit — and treat disease.

Helicobacter pylori (herein *H. pylori*) bacteria infect the stomachs of over one-third of the world's population causing stomach cancer, ulcers, and gastritis (International Agency for Research on Cancer (1994) Schistosomes, Liver Flukes and Helicobacter Pylori (International Agency for Research on Cancer, Lyon, France; <http://www.uicc.ch/ecp/ecp2904.htm>). Moreover, the international Agency for Research on Cancer recently recognized a cause-and-effect relationship between *H. pylori* and gastric adenocarcinoma, classifying the bacterium as a Group I (definite) carcinogen. Preferred antimicrobial compounds of the invention found using screens provided by the invention, particularly broad-spectrum antibiotics, should be useful in the treatment of *H. pylori* infection. Such treatment should decrease the advent of *H. pylori*-induced cancers, such as gastrointestinal carcinoma. Such treatment should also cure gastric ulcers and gastritis.

Vaccines

Another aspect of the invention relates to a method for inducing an immunological response in an individual, particularly a mammal which comprises inoculating the individual with a polypeptide of the invention, or a fragment or variant thereof, adequate to produce antibody and/ or T cell immune response to protect said individual from infection, particularly bacterial infection and most particularly *Streptococcus pneumoniae* infection. Also provided are methods whereby such immunological response slows bacterial replication. Yet another aspect of the invention relates to a method of inducing immunological response in an individual which comprises delivering to such individual a nucleic acid vector to direct expression of a polynucleotide or polypeptide of the invention, or a fragment or a variant thereof, for expressing such polynucleotide or polypeptide, or a fragment or a variant thereof *in vivo* in order to induce an immunological response, such as, to produce antibody and/ or T cell immune response, including, for example, cytokine-producing T cells or cytotoxic T cells, to protect said individual from disease, whether that disease is already established within the individual or not. One way of administering the gene is by accelerating it into the desired cells as a coating on particles or otherwise. Such nucleic acid vector may comprise DNA, RNA, a modified nucleic acid, or a DNA/RNA hybrid.

A further aspect of the invention relates to an immunological composition which, when introduced into an individual capable or having induced within it an immunological response, induces an immunological response in such individual to a polynucleotide of the

invention or protein coded therefrom, wherein the composition comprises a recombinant polynucleotide or protein coded therefrom comprising DNA which codes for and expresses an antigen of said polynucleotide or protein coded therefrom. The immunological response may be used therapeutically or prophylactically and may take the form of antibody immunity or cellular immunity such as that arising from CTL or CD4+ T cells.

A polypeptide of the invention or a fragment thereof may be fused with co-protein which may not by itself produce antibodies, but is capable of stabilizing the first protein and producing a fused protein which will have immunogenic and protective properties. Thus fused recombinant protein, preferably further comprises an antigenic co-protein, such as lipoprotein D from *Hemophilus influenzae*, Glutathione-S-transferase (GST) or beta-galactosidase, relatively large co-proteins which solubilize the protein and facilitate production and purification thereof. Moreover, the co-protein may act as an adjuvant in the sense of providing a generalized stimulation of the immune system. The co-protein may be attached to either the amino or carboxy terminus of the first protein.

Provided by this invention are compositions, particularly vaccine compositions, and methods comprising the polypeptides or polynucleotides of the invention and immunostimulatory DNA sequences, such as those described in Sato, Y. *et al.* Science 273: 352 (1996).

Also, provided by this invention are methods using the described polynucleotide or particular fragments thereof which have been shown to encode non-variable regions of bacterial cell surface proteins in DNA constructs used in such genetic immunization experiments in animal models of infection with *Streptococcus pneumoniae* will be particularly useful for identifying protein epitopes able to provoke a prophylactic or therapeutic immune response. It is believed that this approach will allow for the subsequent preparation of monoclonal antibodies of particular value from the requisite organ of the animal successfully resisting or clearing infection for the development of prophylactic agents or therapeutic treatments of bacterial infection, particularly *Streptococcus pneumoniae* infection, in mammals, particularly humans.

The polypeptide may be used as an antigen for vaccination of a host to produce specific antibodies which protect against invasion of bacteria, for example by blocking adherence of bacteria to damaged tissue. Examples of tissue damage include wounds in skin or connective tissue caused, e.g., by mechanical, chemical or thermal damage or by implantation of indwelling devices, or wounds in the mucous membranes, such as the mouth, mammary glands, urethra or vagina.

The invention also includes a vaccine formulation which comprises an immunogenic recombinant protein of the invention together with a suitable carrier. Since the protein may be broken down in the stomach, it is preferably administered parenterally, including, for example, administration that is subcutaneous, intramuscular, intravenous, or intradermal. Formulations suitable for parenteral administration include aqueous and non-aqueous sterile injection solutions which may contain anti-oxidants, buffers, bacteriostats and solutes which render the formulation isotonic with the bodily fluid, preferably the blood, of the individual; and aqueous and non-aqueous sterile suspensions which may include suspending agents or thickening agents. The formulations may be presented in unit-dose or multi-dose containers, for example, sealed ampules and vials and may be stored in a freeze-dried condition requiring only the addition of the sterile liquid carrier immediately prior to use. The vaccine formulation may also include adjuvant systems for enhancing the immunogenicity of the formulation, such as oil-in water systems and other systems known in the art. The dosage will depend on the specific activity of the vaccine and can be readily determined by routine experimentation.

While the invention has been described with reference to certain protein, such as, for example, those set forth in Table 1, it is to be understood that this covers fragments of the naturally occurring protein and similar proteins with additions, deletions or substitutions which do not substantially affect the immunogenic properties of the recombinant protein.

Compositions, kits and administration

The invention also relates to compositions comprising the polynucleotide or the polypeptides discussed above or their agonists or antagonists. The polypeptides of the invention may be employed in combination with a non-sterile or sterile carrier or carriers for use with cells, tissues or organisms, such as a pharmaceutical carrier suitable for administration to a subject. Such compositions comprise, for instance, a media additive or a therapeutically effective amount of a polypeptide of the invention and a pharmaceutically acceptable carrier or excipient. Such carriers may include, but are not limited to, saline, buffered saline, dextrose, water, glycerol, ethanol and combinations thereof. The formulation should suit the mode of administration. The invention further relates to diagnostic and pharmaceutical packs and kits comprising one or more containers filled with one or more of the ingredients of the aforementioned compositions of the invention.

Polypeptides and other compounds of the invention may be employed alone or in conjunction with other compounds, such as therapeutic compounds.

The pharmaceutical compositions may be administered in any effective, convenient manner including, for instance, administration by topical, oral, anal, vaginal, intravenous, intraperitoneal, intramuscular, subcutaneous, intranasal or intradermal routes among others.

In therapy or as a prophylactic, the active agent may be administered to an individual as an injectable composition, for example as a sterile aqueous dispersion, preferably isotonic.

Alternatively the composition may be formulated for topical application for example in the form of ointments, creams, lotions, eye ointments, eye drops, ear drops, mouthwash, impregnated dressings and sutures and aerosols, and may contain appropriate conventional additives, including, for example, preservatives, solvents to assist drug penetration, and emollients in ointments and creams. Such topical formulations may also contain compatible conventional carriers, for example cream or ointment bases, and ethanol or oleyl alcohol for lotions. Such carriers may constitute from about 1% to about 98% by weight of the formulation; more usually they will constitute up to about 80% by weight of the formulation.

For administration to mammals, and particularly humans, it is expected that the daily dosage level of the active agent will be from 0.01 mg/kg to 10 mg/kg, typically around 1 mg/kg. The physician in any event will determine the actual dosage which will be most suitable for an individual and will vary with the age, weight and response of the particular individual. The above dosages are exemplary of the average case. There can, of course, be individual instances where higher or lower dosage ranges are merited, and such are within the scope of this invention.

In-dwelling devices include surgical implants, prosthetic devices and catheters, i.e., devices that are introduced to the body of an individual and remain in position for an extended time. Such devices include, for example, artificial joints, heart valves, pacemakers, vascular grafts, vascular catheters, cerebrospinal fluid shunts, urinary catheters, continuous ambulatory peritoneal dialysis (CAPD) catheters.

The composition of the invention may be administered by injection to achieve a systemic effect against relevant bacteria shortly before insertion of an in-dwelling device. Treatment may be continued after surgery during the in-body time of the device. In addition, the composition could also be used to broaden perioperative cover for any surgical technique to prevent bacterial wound infections, especially *Streptococcus pneumoniae* wound infections.

Many orthopedic surgeons consider that humans with prosthetic joints should be considered for antibiotic prophylaxis before dental treatment that could produce a bacteremia. Late deep infection is a serious complication sometimes leading to loss of the prosthetic joint and is accompanied by significant morbidity and mortality. It may therefore be possible to extend the use of the active agent as a replacement for prophylactic antibiotics in this situation.

In addition to the therapy described above, the compositions of this invention may be used generally as a wound treatment agent to prevent adhesion of bacteria to matrix proteins exposed in wound tissue and for prophylactic use in dental treatment as an alternative to, or in conjunction with, antibiotic prophylaxis.

Alternatively, the composition of the invention may be used to bathe an indwelling device immediately before insertion. The active agent will preferably be present at a concentration of 1µg/ml to 10mg/ml for bathing of wounds or indwelling devices.

A vaccine composition is conveniently in injectable form. Conventional adjuvants may be employed to enhance the immune response. A suitable unit dose for vaccination is 0.5-5 microgram/kg of antigen, and such dose is preferably administered 1-3 times and with an interval of 1-3 weeks. With the indicated dose range, no adverse toxicological effects will be observed with the compounds of the invention which would preclude their administration to suitable individuals.

Each reference disclosed herein is incorporated by reference herein in its entirety. Any patent application to which this application claims priority is also incorporated by reference herein in its entirety.

TABLES

Certain pertinent data for preferred polypeptide and polynucleotide embodiments of the invention are summarized in Tables 1 and 2.

Provided in Table 1 are sequence search results providing characterization information regarding certain preferred polynucleotides (denoted as "Assembly") and polypeptides of the invention encoded thereby. For each polynucleotide in Table 1, there is listed the closest homologue of each polypeptide encoded by each ORF in such polynucleotide. This determination of homology is based on a comparison of the sequences of in Table 1 with sequences available in the public domain (see heading entitled "Description" for the homologue name). Where no significant homologue was detected the term "unknown" appears after the heading "Description". Preferred polypeptides encoded by the ORFs of the invention, particularly full length proteins either obtained using such

ORFs or encoded entirely by such ORFs, are ones that have a biological function of the —
homologue listed, among other functions. The analysis used to determine each homologue
listed in Table 1 was either BlastP and/or BlastX and/or MPSearch, each of which is well
known. Also provided in Table 1 is the amino acid sequence encoded by each ORF. An
"Assembly ID" number provides a convenient way to correlate the polynucleotide sequence
with the ORF or ORFs it comprises and the polypeptides encoded by these ORFs, as well
as to correlate such sequences with other pertinent information provided in Tables 1 and 2.
Following the heading "ORF Predictions" the nucleotides at the beginning and end of the
ORF sequence are set forth ("Start" and "End" respectively). The direction of translation on
the polynucleotide depicted is denoted by an "F" for forward or an "R" for reverse (reverse
being translated on the opposite strand from the one depicted). The length of each amino
acid sequence is also indicated in a column entitled "Length." Below these data is shown
the amino acid sequence encoded by the ORF. If a given polynucleotide comprises one
ORF, then in the column entitled "ORF #" there is the numeral one. If it encodes two, there
are the numerals one and two in the column, and so on.

TABLE 1

Assembly ID: 3049156
Assembly Length: 495bp

```
> 3049156 Strep Assembly -- Assembly id#3049156
CTCGGTGATAGAAATAGTGTAATCATGCTTTTCTCTTCTTATCTATACTTTGCTACTTCT
ATTATACAAAAAATAAAGCGCTTGACTAGGGATTTTGTAGAAAAAAGCCTATTTTTTCA
AGAAAAATAGGCTTTTTTGCGAACGATTGACACAATTGGATTGGTTAATTCACCTCTTAAC
GATGGTTTTAAACGATATATATTTTTATATATGTAAATTAAAACTTCTTTCTTTCACT
TCCTACGACTTTTCAGATACAGATAGCCAAAGAAGTTTTCATAGAGGGCAAAAAAGAGGA
GGAAGGCATGAAGAAAGAAGGTCTCTGGCAAAATCATAATAACAGGATCCTTGGCTGGAT
CAAAAAGCCAGGTATCATCTCCCAAAAGAGAATTTGATGGAAAAGAGTAAAGAATTGGT
CAAAACCAATCAAACTCCCCCAAGTCCATCATCACAGGTAAGACTACTAGAGCCAGGAG
ACTTTTTTCGATAAAG
```

ORF Predictions:

ORF #	Start	End	Direction	Length
-----	-----	-----	-----	-----
1	236	385	R	50 aa

```
> 3049156-1 ORF translation from 236-385, direction R
VGDDTWLFDPAKDPVIMILPETFFLHAFLLFFALYENFFGYLYLKSRK*
```


Description:
unknown

Assembly ID: 3049862
Assembly Length: 529bp

```
> 3049862 Strep Assembly -- Assembly id#3049862
CTAGAGCAAGTATTTTCAAACCTTTTCCGAATAAATAGATAGAGCCAGAGAATTTAGTA
AACCTAGATTTTAAAAATGTGCTATAACATAATATATTGAATCTATAATAGTACACCTTGA
CTGCTAAAATATTTCTATAAAATTAATTTGACTTTCCTGATAGAGTTATTCACATCTTATT
TCAACTCACTATAGAAGGAGGAATAGGAGGATTCTCAGACATCCGGGCATCAGCCCAACT
AATGATTTGATTGCTAAGAAAATATTCAGCAATCCAGAAATCACTTGTCAATTTATTCGC
GATATGCTGGACTTGCCAGCAAAAAATGTTGACCATTTTGGAGGGAAGCGATATTCACGT
ATTACTCTCCATGCCTTACTCAGTGCAGGATTTTTTATACCAGTATAGACGTCTTGGCGGA
GTTGGATAACGGTACTCAAGTAATTATTGAGATTCAAGTCCATCATCAGAATTTTTCATC
AATCACTTGTGGACTTACCTGTGCAGTCAGGTTAATCAAATCTTGAAAA
```

ORF Predictions:

ORF #	Start	End	Direction	Length
1	383	526	F	48 aa

```
> 3049862-1 ORF translation from 383-526, direction F
VQDFYTSIDVLAELDNGTQVIIEIQVHHQNFSSITCGLTCAVRLIKS*
```

Description:
unknown

Assembly ID: 3112810
Assembly Length: 885bp

```
> 3112810 Strep Assembly -- Assembly id#3112810
CTCATCATCTGTCAAAAAGCGTTTCTTAGCAGTCGTGATATCCATAAAATAATCTAATAT
CACGATTTCTTCATCCGCAAAGAAAGGAAGGCTGACCAACTCCAGTGCCACATCCTTGTA
AACTACTTCTTGCATATCAAAGTAGGCAAAGTTGAGGTCAGCAGAATCATACCCAATCTG
TTTCAACACTTGACTCTTCATCACTTCAAACCTGACCCCTGATCTGTCCCTGTAAATAGGCG
CAGGCTCGGTAAATTCGATAAAGTCAACTTCTGACTTTCTTCAATGGCTAGCATCGTCTC
TCCTTTCTTCAGATTTTTCGATTTAATTTAGTCAATATAGCGCAATTTCCACGGAAATC
TTCTAAGCTCTCGTAGCCTTTTCCACCATGATTGCTTTCAGTTCATTGGTAAAGCGGTC
AAAAGCACTGACGCCTTCTTTGTGAAGGGTCGTTCCACCTGCACCATACTTGCTCCACA
GAGGATGTGTTCAAAGGCATCTCGACCAGTCAGAACGCCACCTGTTCCGATAATTTGGAT
TTGAGGATTTAAACGTTGATAAAAGGCGTGAACATTGGCTAGAGCAGTCGGTTTGATGTA
```

TTATCCACCAATTCCACCAAACCATTCCTTAGGCCGAATAACGACAGATTCGTCTTCTAT
 ATAGAGGCCGTTTCCGATAGAGTTAACGCAGTTGACAAACTTGAGCGGATATTTGTTGAA
 AATAGCTGCCGCTTGATCAAAGTGAACAATATCAAAATAAGGTGGCAATTTAATTCCAAG
 AGGTTTGGTGAAGTAAGCAAACACTTCTGCCAAAATCCGGTCTGTTGTCTCAAAATCATA
 GGCAATCTGAGGTTTACCTGGAACATTTGGACAGGAAAGATTTAG

ORF Predictions:

ORF #	Start	End	Direction	Length
1	601	804	R	68 aa

> 3112810-2 ORF translation from 601-804, direction R

VFAYFTKPLGIKLPPYFDIVHFDQAAAIIFNKYPLKFVNCVNSIGNGLYIEDESVMIRPKN
 GFGGIGG*

Description:

LLCPYRDA NCBI gi: 511014 - Lactococcus lactis. DIHYDROOROTATE
 DEHYDROGENASE (EC 1.3.3.1) (DIHYDROOROTATE OXIDASE)

Assembly ID: 3112866

Assembly Length: 925bp

> 3112866 Strep Assembly -- Assembly id#3112866

TCTTGGCCAACTGCATGGAGTTCAGCGGTCAATTTCAACGCACCTGAGAAACAGACCCCT
 GCACCCCTGAAATCTCAGGAGACATGATGGTCTGGATGGAATCAATAATGAGAAAGTCTG
 GCTGGATACGCTACCACTTCTGCACGAACACTCTGCATATTGGTCTCTGCATAGAGATAA
 AACTCACTATCAAAATCACCTAAGCGCTCTGCACGTAGTTTAATCTGCTGGGCAGACTCC
 TCCCCACTGACATAGAGAACTGTCCCCACTTGGGACAACCTGGGTTGAGACTTGTAGGAGA
 AGAGTTGATTTCCCAATCCCAGGATCCCCACCGATGAGGACGAGACTTTCCTGGTACAAC
 TCCGCCTCCAAGCACACGGTTGAATTCCTCCATCTCCGTCTTGGTTTCGATTGACATTGAT
 GGAAGTCACCTCAGCTAGTTTTCATGGGCTTGGTTTTCTCACCTGTCAAGGACACACGCGC
 ATTCTTGACCTCGGCAACCTCAACCTCTTCCACAAAAGAAGACCAAGACCCACAGTTGGG
 GCAACGTCCCAGATATTTAGGGGAATTATACCCACAATTTTGACATACAAATGTCGCTTT
 TTTCTTTGCGATGACAAACCTCTTTCTATATCTCTAACTCACACTCAATCACTTGGCAAA
 AATCAATCTTCTCATTTGGCACAACCTGGCGCATGAGCATTCGATGAGCAACAACCTACCA
 CAGTCTGATGTTCTCGATACTTAGACATACATTCTAGAAACCGAGACTTCATTTCCGTAG
 CTGTCTCATATTGAATAGGACTATTAGGAAGCAACTCCCCCTTGTTTTCTAAAAACAGTC
 TTCTAGCTGTTTCAAAGTTTTCTATTCCTGTTTTATAGACCTGCCATTCATGTAATAAAG
 GCTCTACTCTTAAAGGAAGACCCGT

ORF Predictions:

ORF #	Start	End	Direction	Length
-------	-------	-----	-----------	--------

1 220 513 R 98 aa

> 3112866-2 ORF translation from 220-513, direction R
VEEVEVAEVKNARVSLTGEKTKPMKLAEVTSINVNRTKTEMEEFNRLGGGVVPGKSRPH
RWGSWDWEINSSPTSLNPVVPSPGDSSSLCQWGGVCPAD*

Description:

SMS PROTEIN. - ESCHERICHIA COLI.

Assembly ID: 3113664

Assembly Length: 602bp

> 3113664 Strep Assembly -- Assembly id#3113664

TTATGTCAGTGGGATTACGCCTAATCTCCCAGAAGCAGAATTATTATCCGGTCAGGAAAT
TAAACCTTGGNAGACATGAAACTGCAGCGCAGAAATTGCATGATTTAGGAGCGCCAGC
AGTCATTATCAAAGGGAGGCAATCGTCTTAGTCAGGACAAGGCTGTGGATGTCTTTTATG
ATGGACAGACCTTTACTATCCTAGAAAATCCAGTTATCCAAGGCCAAAATGCTGGTGCAG
GTTGTACCTTTGCCTCTAGCATTTGCCAGTCACTTGGTTAAAGGTGATAAACTTTTGCCAG
CAGTAGAAAGCTCTAAGGCTTTCGTTTATCGTGCTATTGCACAAGCAGATCAGTATGGAG
TAAGACAATATGAAGCAAACAAAAACAATAAAATCGCCCTTGTATCCCTATTAACCGCC
CTTTCTGTGGTTCTAGGTTATTTCTTAAAAATCCCAACACCTACAGGNATTCTAACTCTT
TTAGATGCTGGTGTCTTCTTTGCGGCCTTTTACTTTGGTAGTCGTGAAGGAGCGGTAGTC
GGAGGACTAGCAAGTTTCTTGCTTGACCTCTTATCAGGCTACCCTCAGTGGATGTTTTTT
AG

ORF Predictions:

ORF #	Start	End	Direction	Length
1	165	392	F	76 aa

> 3113664-1 ORF translation from 165-392, direction F
VDVFYDGQTFITILENPVIQGNAGAGCTFASSIASHLVKGDKLLPAVESSKAFVYRAIAQ
ADQYGVRQYEANKNN*

Description:

Thi protein - Rhizobium meliloti

Assembly ID: 3113716

Assembly Length: 456bp

> 3113716 Strep Assembly -- Assembly id#3113716

CTGGATACTAAGAGAAATCAAAAAAGCACTCTAGGATAGAGGCCTAAAGTGCTTAGTTTC
AAGGCTTTACAGCCTATCATATTTAATAAAATATTACAACATCTTGTTGTAGAATTCAAC

GACAAGTGCTTCGTTGATTTCTGGGTTGATTTTCGTCGCGTTCTGGCAAGCGAGTCAATGA
 ACCTTCCAATTTTTCAGCGTCGAATGATACGAATGCTGGACGTCCAAGAGTAGCTTCTAC
 TGCTTCAAGGATTGCTGGAACCTTCAATGATTTTTCACGAACTGAGATCACTTGACCTGC
 AGTTACGCGGTATGATGGGATATCAACGCGTTTCCCGTCAACAAGGATGTGACCGCTGGT
 TTACAAATTGGACCAAACCTTGACGACCAGTAGTCGCGAGACCAAGACGGTAAACAACGTT
 ATCCAAACGACGTTCCAAAAGAAGCATAAAGTTGAA

ORF Predictions:

ORF #	Start	End	Direction	Length
1	94	291	R	66 aa

> 3113716-1 ORF translation from 94-291, direction R
 VISVREKSLKVPAILAEVEATLGRPAFVSFDAEKLEGS LTRLPERDEINPEINEALVVEF
 YNKML*

Description:

30S RIBOSOMAL PROTEIN S4 (BS4). - BACILLUS SUBTILIS.

Assembly ID: 3174176

Assembly Length: 1961bp

> 3174176 Strep Assembly -- Assembly id#3174176

CTAATATAGAATAATCACCGCCGTTGTGAAAGAACGATTGGATGATAATCCAATCGTTCA
 GGGAAATTGGAAGACCTTGGGTTTCCAATTTAGGCATGAGACACCTTTGGTGGCTGCTGC
 CGTCCCTCACAAGCTAAGGTGATTGTTGAAAAAGAGGAAAAAGGAGAAGAAATGAAACCA
 GTAATTTCCATCATCATGGGCTCAAAATCCGACTGGGCAACCATGCAAAAAACAGCAGAA
 GTCCTAGACCGCTTCGGTGTAGCCTACGAAAAGAAAGTTGTTTCCGCACACCGTACACCA
 GACCTCATGTTCAAACATGCAGAAGAAGCCCGTAGTCGTGGCATCAAGATCATCATCGCA
 GGTGCTGGTGGCGCAGCGCATTTGCCAGGCATGGTAGCTGCCAAAACAACCTTCCAGTC
 ATTGGTGTGCCAGTCAAGTCTCGTGCTCTTAGTGGAGTGGATTCACTCTATTCTATCGTT
 CAGATGCCGGGTGGGGTGCCTGTTGCGACCATGGCTATCGGTGAACTCTTTTTTAGGATA
 TAAACAGGGTTCGGATAAGTTTTTTTGCAAGGTGGATGATGGCTACATTGTAATGTTTTT
 CCTTGTTCTAACTTAGTCTTAAAAGCAGGTGAAAAGTGAGGGCATGCTTTGGCAGCTTGT
 ATGAGTACCTACCGCAGATAAGGGGAACCCCGTTTGACCATCCTCCCAGCTAAATCAATC
 TGACCTGACTGATAAATAGAAGAATCCAGTCCAGCGAAAGCTTGTAATTGAGCAGGATTA
 TCAAAGGCATGAATATTTTCAATCTCGGCTAAATGACCGCCCCCTAAACGATTCTCAATC
 CCAGTAACCGTCGTGATGACCGAGTTTAACTCAGCCATCAAGTCATTGACACATTTTTCC
 GCCTTGTCATGAGCCTCTTGTAATGTTTGATGTTTTTCATTACACGAGATAAAACGTCTA
 TGCGTTATCAAATCATTAACCAATTAACAAATGTGGTTAGATCCTTTTCGGAAATTGTC
 AAGCGATTGGAGGAAATGAACTAATCCACAGCGGCTTATTTCCAAGTATACCACTTGGGCT
 TTGGCAGTAGCTAACTGCGCTAAATATAATATAAGGAGGAGTAAAATGAAGACAGTTCAA
 TTTTTTTGGCATTATTTTAAGGTCTACAAGTTCTCATTGTAGTTGTCATCCTGATGATT

GTTCTGGCGACTTTTGCCCAAGCCCTCTTTCCAGTCTTTTCTGGACAAGCGGTGACGCAG
 CTAGCCAATTTAGTTCAAGCTTATCAAAATGGGCAATCCAGAACTTGTATGGCAAAGCCT
 ATCAGGAATTCATGGTCAATCTTGGCCTGCTGGTTTTGGGTTCTATTTATCTCTAGGTGT
 AATATAAACATGTGTCTCATGACGCGCGTGATTGCAGAATCGACCAACGAGATGCGCAA
 GGTCTCTTTGGTAAGCTTGCTCAGTTGACGGTTTCTTTCTTTGACCGTCGACAAGATGGC
 GATATCCTGTCTCATTTTACCAGTGATTTGGATAATATCCTCCAAGCCTTTAACGAAAGC
 TTGATTCAGGTCATGAGCAATATTGTTTTATACATTGGTCTGATTCTTGTCTATGTTTTCG
 AGAAATGTGACGCTGGCTCTCATCACCATTGCCAGCACCCCATTTGGCTTTCCTTATGCTG
 ATTTTCATCGTGAAAATGGCACGTAAATACACCAACCTCCAGCAGAAAGAGGTAGGGAAG
 CTCAACGCCTATATGGATGAGAGCATCTCAGGCCAAAAAGCCGTGATTGTGCTAGGAATT
 CAAGAGGATATGATGGCAGGATTTCTTGAACAAAATGAGCGCGTGCGCAAGGCAACCTTT
 AAAGGAAGAATGTTCTCAGGAATTCTTTTCCCTGTCATGAATGGGATGAGCCTGATTAAT
 ACAGCCATCGTCATCTTTGCTGGTTCGGCTGTACTTTTGAA

ORF Predictions:

ORF #	Start	End	Direction	Length
1	139	543	F	135 aa

> 3174176-1 ORF translation from 139-543, direction F

VIVEKEEKGEEMKPVISIIMGSKSDWATMQKTAEVLD RFGVAYEKKV VSAHRTPDLMFKH
 AEEARSRGIKIIIIAGAGGAHLPGMVAAKTTL PVIGVPVKSRALSGVDSLVSIVQMPGGV
 PVATMAIGELFFRI*

Description:

PHOSPHORIBOSYLAMINOIMIDAZOLE CARBOXYLASE CATALYTIC SUBUNIT (EC
 4.1.1.21) (AIR C ARBOXYLASE) (AIRC). - BACILLUS SUBTILIS.

Assembly ID: 3174186

Assembly Length: 375bp

> 3174186 Strep Assembly -- Assembly id#3174186

CTATCTCCAAGTNCGNNTTGAATNCCTCCGCNANCCACAACCTCATCCAAGCACTTTNCAA
 CGTGNCCTGGTCCGGTCCCTCCAGTGCGTCTNACNGCACCTTCAACCTGCNCATGGGTAGG
 TCACATGGCTTCGGGTCTACGTCATGATACTAAGGCGCCCTATTCAGACTCGGNTNCCCT
 AGGGCTCCGTCTCTTCAACTTAACCACGCAACAGAACGTNACCCGCCGGTTCATTCTACA
 AAAGGCAGNCTCTCACCCATTAACGGGCTCGAACTTGTTGTAGGCACACNGCTTCAGGTN
 CTATTTTACCCCCCTCCCGGGGAGCANCTCAACTGACCCNCACGGCACCGGTGNANNAAA
 CGGTCACTTAGGGAG

ORF Predictions:

ORF #	Start	End	Direction	Length

1 83 283 F 67 aa —

> 3174186-1 ORF translation from 83-283, direction F
VRXXAPSTCXWVGHMASGLRHDTKAPYSDSXXLGLRLFNLTQQNXTRRFILQKAXSHPL
TGSNLL*

Description:
unknown

Assembly ID: 3174374
Assembly Length: 665bp

> 3174374 Strep Assembly -- Assembly id#3174374
GGGGGGGGTNNNTTCTGGGGCCGGGTGNNTCCTNGAAAAAATGCTGGACTTAACGGTTAA
ATCATTTGAATTGGCCTGTGGATTTTCTAGCTAGCAATCCAGAGCGAGTTTCTCCAAGACA
GACCTCTATGAAAAGATCTGGAAAGAANACTACGTGGATGACACCAATACCTTGAATGTG
CATATCCATGCTCTTCGACAGGAGCTGGCAAAATATAGTAGTGACCAAACGCCCACTATT
AAGACAGTTTGGGGGTTGGGATATAAGATAGAGAAACCGAGAGGACAAACATGAAACTAA
AAAGTTATATTTTGGTTGGATATATTATTTCAACCCTCTTAACCATTTTGGTTGTTTTTT
GGGCTGTTCAAAAAATGCTGATTGCGAAAGGCGAGATTTACTTTTTGCTTGGGATGACCA
TCGTTGCCAGCCTTGTCGGTGCTGGGATTAGTCTCTTTCTCCTATTGCCAGTCTTTACGT
CGTTGGGCAAACTCAAGGAGCATGCCAAGCGGGTAGCGGCCAAGGATTTCCCTCCAATTT
GGANGTTCAAGGTCCCTGTAAATTTCCCCCATTTAGGGGCAACCTTTTAATGAAANTTT
CCNTNATTTGCCGGGTANCTTTGAATCCCTNGGAAAAAACCCAACNAAAAAAGGGCTTA
NNCCC

ORF Predictions:

ORF #	Start	End	Direction	Length
1	154	294	F	47 aa

> 3174374-1 ORF translation from 154-294, direction F
VDDTNTLNVHIALRQELAKYSSDQTPTIKTVWGLGYKIEKPRGQT*

Description:
REGULATORY PROTEIN VANR. - ENTEROCOCCUS FAECIUM (STREPTOCOCCUS
FAECIUM) .

Assembly ID: 3174972
Assembly Length: 989bp

> 3174972 Strep Assembly -- Assembly id#3174972
CTACGATATCTTTGGTCTTTTGTAAAGATATGAGGTCCACCCTTATGCGCCTCAGTTGGCA

TTTCATGCGATTCAAGAAGTTGCCCCCTCTTGATCAACCAAACCATACTTGATGTTGGTTC
 CACCGATATCAATTGCAACGTAATATGTCATAAATACCTCCTTTTAGATTAGAGGAAGCG
 CTCCTTGGTTTCACGAATCAAGGCAGCAGCCGCTTCTACAACCTGGACGATCTTCTTCAGT
 CACTGGTGTCAATGGTGAACGAACAGATCCAATATTCAAGCCTTCATTGATTTTCAAGAC
 TTCTTTGATGACACCGTACATATTTCCATGAGCAGAAAGTGAGTTTACCAATGATTGCGTT
 GATAGCATACTGCAATTCACGCGCTGTTTCTAGGTCCTTATCCGCAATCAACTGATTGAG
 TTTCAAGAAGAGTTCTGGCATAGCACCATAAGTACCACCGATACCAGCCCTAGCCCCCAT
 GAGGCGTCCTCCTAGGAAGTCTCATCAGGACCATTAAAGACGATATGGTCTTCTCCACC
 AAGGCTGACAAAGGTTTGGATATCTTGAAGTGGCATAGAAGAGTTCTTCACACCGATAAC
 ACGAGGATTTTTCAACATTTCTGTGTAAAGGCTTGGAGTCAAAGCAACCCCTGCCAATTG
 AGGAATGTTGTAAATCACGTAGTCTGTGTTTGGAGCTGCAGAACTGATATCGTTCCAGTA
 TTTGGCAACTGAGTTATTCTGGCAAGCGGAAATAAATTGGTGGAATCCGTTGCAATAGCA
 TCTACTCCCAAGCTTTCAGCATGGCGAGCAAGTTCCATACTATCTTTAGTATTATTGCAA
 GCAACATGGGCAATAATGGTCAATTTACCTTTGGCTACCGCCATGACTTCTTCCAAAATC
 AACTTGCGATCTTCAACGCTTTGGTAGATACATTCACCAGAAGAACCATTGACATAAGAC
 CTTGAACACCTTTATCAATGAAGTATTGA

ORF Predictions:

ORF #	Start	End	Direction	Length
1	169	678	R	170 aa

> 3174972-1 ORF translation from 169-678, direction R
 VIYNIPQLAGVALTPSLYTEMLKNPRVIGVKNSSMPVQDIQTFVSLGGEDHIVFNGPDEQ
 FLGGRLMGARAGIGGTYGAMPELFLKLNQLIADKDLTARELQYAINAIIGKL TSAHGM
 YGVIKEVLKINEGLNIGSVRSPLTPVTEEDRPVVEAAAALIRETKERFL*

Description:

N-ACETYLNEURAMINATE LYASE SUBUNIT (EC 4.1.3.3) (N-
 ACETYLNEURAMINIC ACID ALDOLAS E) (N-ACETYLNEURAMINATE PYRUVATE
 LYASE) (NALASE). - ESCHERICHIA COLI.

Assembly ID: 3175138

Assembly Length: 1450bp

> 3175138 Strep Assembly -- Assembly id#3175138

CTCCATATTTCTTAGCCTTCTCAATTAGGGTCTTGAAGTCTTCGACACCACCGATACGCT
 TACCAATATCAGCATAGTTCAAGTGACCAGAGTCATGGCTGTGATATCCTTAACCTTTTTC
 CCAACCTTGAGGGTTGTTTCATAATGCTACGATAAGCAATGGCACCATCTTGCCAATCAAC
 TTTCTTGTCTGCATTGGCATCTTCAGTGATAACAACCTTAGCACTTGGAAGTTCCTTCGT
 GTATTCTGGGAAAACAATGCCCTTATAAGCTTTTTCCCATTTGCCATTCAGAGCTGTGGAT
 TCCTACATAGTTGGCATTTCCGACTGTTTCTTTTATAAGCTGTCAAACGAGTCCAGTCATT
 CGAACCACCACCATAGCTATTTTGAGAGTTACTCCAAACACCAGCAGCAAGCTTATCTGT

AGAAACAAATCCATACATGTAACCCTTAGCCAAATCCTTCATTGGATTGGTTACATCGAT
 ATGATCATCTCCGCTGACATGCGTATTGTTTGACATGGTTGCCCCATCAAACCTTAGCACC
 AGTTTGATCACTAGAAACAGAGACTAAAGCATTGCCGAGGAACTAATAGAAGAAAGTAG
 TTTTCTTTTCGTCATCAATCTTTTGACCTGGAGTGACTTGATTGTGGTTGACAATCTTGGT
 CACATCAAAGTGCAATTGATTGTCCACAACCTTGCAAGCGTACTGTCAATTTCCGCATTGAT
 TAAGTGAGCATCATCGCGAAGCTTCATCAAGTACTCTGCTGTTGTCTCATTGATTTTTTTT
 ATAAGTGACTTCAGGGGTGATTCGGTGGTTATTGATAAAGACTTGGTTGAATTGTTGCAC
 CTGTCCTGGCAAAGTATGTCCATTCAAGGTGTATCCCTTGACACGAAGGAAGGCTTGGTC
 AATTACTGCCTTAAGTACCTTAAACTGGATCGTATCATAAGTCACCTTGCTATCGTCAAC
 AACCGGACCTGTTTCTTTCTGTTGGCAGGGGTATCCTCTGGGTTTTACCCTCTCTGTGGCTA
 TCCGTTTTCAACGCTTGAACAACCTGGTCGCTCATCGTCATAAGAGCCCGCCTTGAGAAAAA
 TCTTCTTCTCATTCTTAAGATGGTCATTGACCGCAGCTGGTAGAGTCACTGTGTCAAAGA
 AGATTGACATCCTTATTTGCCTGGCATTACCTGACCGTCTGACTTGAAGACTGATAGAG
 AGACGGTTTGTGATCCTGTTTCAGGAGCAGCAACACGACTACCTCTATACCAAGTGCTA
 GTTGTGGAGATTTATACTCCCAGAACCAGCCATCCTTGTCTATAACCGACAAAAACATTA
 TTATTGGTATCTTTAAATTTCAAGGAGACACCAAAGCGTGATTTGCCCTTTTCAGAATCT
 TCTTTGAAGGTTAAATCAACAGTTGCATTTCCATTGGCATCAACGGTCAAGCCCTTCTTT
 TCAAACAGAG

ORF Predictions:

ORF #	Start	End	Direction	Length
1	79	945	R	289 aa

> 3175138-1 ORF translation from 79-945, direction R
 VTYDTIQFKVLKAVIDQAFLRVKGYTLNGHTLPQVQQFNQVFINNHRITPEVITYKKINE
 TTAEYLMKLRDDAHLINAEMTVRLQVVDNQLHFDVTKIVNHNQVTPGQKIDDERKLLSSI
 SFLGNALVSVSSDQTGAKFDGATMSNNTHVSGDDHIDVTNPMKDLAKGYMYGFVSTDKLA
 AGVWSNSQNSYGGGSNDWTRLTAYKETVGNANYVGIHSSEWQWEKAYKGIVFPEYTKELP
 SAKVVITEDANADKKVDWQDGAIAYSIMNNPQGWEKVKDITAMTLVT*

Description:

unknown

Assembly ID: 3175860

Assembly Length: 420bp

> 3175860 Strep Assembly -- Assembly id#3175860
 CTGCGAGTTGTGAGGCTCCTATTATGTCTCGTGATTAAATCTCTATAAGGTGATTTTGG
 AGGGAAATTATCGGGCGACAGCGGGTAGAGAAGAGATGAAAGAGGCTATTTTGAATATC
 AAGCAAATCCTGCTGCCTTAAAGATCTCAAAGAAAAGGCTAAGAATATTTCCAGAGAGT
 ATTCTGAAGAGCATCTGTTACAAATCTGGTTGGACTTTTATGAGAAACAAGCCGCTTTAG
 GGACAAAGTAAAAAGTGAGGTAATCTATGCGAATTGGTTTATTTACAGATACCTATTTTC

CTCAGGTTTCTGGTGTGCGACCAATATCCCAACCTTGAAAACCCACCTTGAAAACACGG
 ACTTGCCTGCATTTNTATCTCATACAATCCACCGAATTTTCGATGTCCCCCTCCCTACAAC

ORF Predictions:

ORF #	Start	End	Direction	Length
1	51	251	F	67 aa

> 3175860-1 ORF translation from 51-251, direction F
 VILEGNYRATAGREEMKEAILEYQANPAALKDLKEKAKNISREYSEEHLQLIWLDFYEKQ
 AALGTK*

Description:

unknown

Assembly ID: 3175918

Assembly Length: 661bp

> 3175918 Strep Assembly -- Assembly id#3175918

CTCCCCAACTTTTATTTGAGAGTGAACGGTATAAGAATATGAAACCGGAGGTTAAGGTG
 GTTTACTCAGTTTTTAAAAGATCGGTTGGAGTTGTCTTTGAGCAAAGGTTGGATTGATGAG
 GATGGGACTATTTATTTGATTTATTCCAATTCAAATTTGATGGCACTTTTAGGCTGTTCA
 AAGTCAAAATTACTCTCCATGTGAGTTTGAAGTGACATTTTATAGATGATTACCATAAAAA
 ACATAACTACCCACTATTTTACGAATCCTATCTTCAAAACGTTATGGAATTCCTTGAAAG
 TCAAGACATAAAGAATGGGGTTGATGCCTTTGTAGATGATCATCAAAATCTCGTTTTTGT
 TTTATATGGACAAGGCTATCGAGCCGAGGGAAAAGAGGGAATACTTACAACCCAAGTAAC
 TGTAAGGCTTATGATGAAGACAAGAAACCGATTAACCTTCGCAAATTTATTAGATTCCTT
 AATCGTGTGAGAATATCAAATGGAACCGAATCTTTGGGAGGTCTCCTATGATTGATCTCT
 ATCTAAGTAAAAATAGCCGAAGAAATCAACTTCTTTTAGACTTCTTCCAAACTATGGCA
 TCGAGGTATCTTGTCAATTCAGTTTCTGAAATGACAAAGGACAAATTAATTGAGATGATGA
 G

ORF Predictions:

ORF #	Start	End	Direction	Length
1	212	535	F	108 aa

> 3175918-1 ORF translation from 212-535, direction F
 VTFLDDYHKKHNYPLFYESYLQNVMEFLESQDIKNGVDADFVDDHQNLVFLYGGQGYRAEG
 KEGILTQVTVKAYDEDEKKPINFANLLDSLIVSEYQMEPNLWEVSYD*

Description:

unknown

Assembly ID: 3811220
 Assembly Length: 1429bp

> 3811220 Strep Assembly -- Assembly id#3811220

CTGCCCCCTGTAAGGCTGGACGATTGCCTTTCTTAGTATCCGCAAAGAGGTAAACTGAGAA
 TAGAGAGGATTTCTCCTTCAATATCTTTGACAGACAGGTTTCATCTTGCCTTCTACGTCTG
 AAAAAATCCGCATATTGACCAGTTTTTCTCACAGCATAGTCCAAATCTTCCTCTTGGTCCT
 CTGGTCCAACACCAACCAGCAATAAAAGTCCCTGATTGATTTTTCCCTGAATCTGGCCTT
 CTATACTCACTTGGGCTTTTTTAACCCGTTGGATAATGATTTTCATAATAGCCTTTCTAG
 TAAGAGCTAGGACAACCTAGCCGTTGGTCCGTTTGACAGAGTAAACTTCTGGCACACTCTT
 AATTTTATCGACAACCGTGGTCAGTGTAGAGAGGTTGGCAATACCGAAGGACACATGGAT
 ATTAGCAAACCTTCATATCCTTGGTTGGTTGGGCATTGACCGTTGAAATATTCTTGGTTGT
 ATTTGAAAGAACTTGCAGTACATCGTTCAACAGTCCGTGTACGGTTGAGACCGTAGATATC
 GATATGGGCCATATACTCCTTATTTGAGCTAGAGTACTGGTCTTCCCATTCCACATCAAG
 GAGACGTTGCTCGTAGTTTTCTTGGGCACGCAGGTTTCATACAGTCCACACGGTGAATAGC
 CACACCACGACCCTTGGTAATGTAGCCAACAATATCGTCACCAGGCACGGGGTTACAACA
 CTTAGCAATCCGCACTAGGAGACCAGAAGCACCTTCAATAACCACTCCCCCTCATGCTT
 GACCTTGGAGAGTTTCTTTATTTTCAACCTTGACCTCGCCACCTTTGACAAGCTCCTCTG
 CCTCAGCCTTGGCCTTGGCACGCTCTTCCTCACGGCGTTCTTTTTTCAGTCAGACGGTTAA
 AGACGGTAATCGCACCGATTTCCCCAAAACCAATGGCCGCAAAGAGGGAGTCTTCTGTCT
 TGTAACCTGGTCTTTTGCAGAACTTGATCCATGTGGCGCTTGTCCATAAATTTATTTGCCA
 CATAGCCATTTTCTTGGAACCTGAGCCATCAGCATCTCACGACCTTGTGACAGACAATT
 CCTTATCTTGGTTTTTTAAAGAACTGGCGAATCTTATTGCGCGCCTTGCTAGTCTTGACCA
 TATTGAGCCAGTCACGGCTAGGTCCAAAGGAGTTCGGGTTGGCGATAATTTCAACCTGAT
 CCCCTGTCTTTAACTTGGTTGTCAGTGGAAACCATGCGGCCATTGACCTTGGCACCAGTTG
 CTTTTTCACCGACCTTGGTATGGATTTTCGTAGGCAAAATCAATCGGTCTGAATCTTTGG
 GAAGAGAACGGACAGCTCCATCTGGGGTAAAAACGTAAATCTCCTCAGCCAGATAGTTTT
 CCTTAACAGAGTCCACAAATTCTTTAGCATCATCAGCCTGGTCTTGGAG

ORF Predictions:

ORF #	Start	End	Direction	Length
1	316	873	R	186 aa

> 3811220-2 ORF translation from 316-873, direction R

VRKSVPRPRLRQRSLSKVARSLKIKKLSKVKHEGGVIEGASGLLVRIAKCCNPVPGDD
 IVGYITKGRGVAIHRVDCMNLRAQENYEQRLLDVEWEDQYSSSNKEYMAHIDIYGLNRTG
 LLNDVLQVLSNTTKNISTVNAQPTKDMKFANIHVSFGIANLSTLTTVVDKIKSVPEVYSV
 KRTNG*

Description:

stringent response-like protein - *Streptococcus equisimilis*

Assembly ID: 3811436
 Assembly Length: 1513bp

> 3811436 Strep Assembly -- Assembly id#3811436

```
CTCTGCAATGATGTACTCAAACATCTCCGCTTCTAGTTCCTCCTTAGGCAGAGGCAATTT
CCCACGTCGCATCCGGTTCATAAAGACCGTATGGTTTTCTAAAATCAAACATATAAACT
CATGTGGGGAATATCCAATCCAATGGCTTTAGCCACATTTTCCTTTACTTGCTCCATGGT
CTGACCAGGCAGAGCATAAATCAAATCAATGGAGATGTTGTCAAACCAGCCAGTTTCAG
GCGATCGATATTTTCATAAATATCCTTCTCCAAATGACTGCGCCCAATCTTTTTCAACAT
CTTATCATCAAAGGTCTGGACACCTAGCGAAACACGATTGACAGCCGAATTTTTCAAAC
AGCTATCTTATCCGCATCCAAATCGCCTGGATTGGCTTCAATGGTCAACTCTTCCAAGAC
AGACAAATCCAAGTTTTTTAGTCAAGCCATTCAGTAACACCTCCAGTTGCGGAGCCGACAG
GGCTGTCTGGTGTTCACCACCGATATAAAGGGTTGACAACTTTTCAATATCATAAGAACG
AAACTCTTCCAGCAGATGCTCTAAATAGCTGTCTGACTGGCTGATTTTTTGATGAAGACCTT
TGAAAAATCACAATAATAACAAATCTGGGTACAAAATGGGATGTGCACATAGGCTGACGT
TGGTTTTTTCTGCATAGTAATTATTATACCACAAAGACTAGATTCCAGATAAAAAATCACC
ATCCCAGATACATAGTCCGTCCGGAGATGGTGATGGTTTATTCTTCTGTTATATCAATC
ACAATCTCTTCTGAGTCATCAAGAGCTTCGGCTTTTTCTTGCCATTGTTTCCTTGAGATTA
TTTAATTGATTTTTTTGATGCTTCTGTCTGCTTGAAAAGCATAGGATTTAGCTTGAGCAAGT
ATACTGTCCACAGTGATTTTACCTGACTCAACCTGTTCTTTTGTTTTTCAGAACAAAATCT
GTAGCCTGCTCCTTAACTTCTGTCTGCTTTTTCACAGACTTGCTCCTTGGCATACTCCGGA
TCTTCTCTCAAATCATCTAAAAAATCTTGAGCCTGACTGCAAACCTGTTTGCCCTTATCA
CTTGTTAAAAACAAGGCAAGAGCTGCACCTGAAACGGTTCCTAAAAGGATTGAGGATAAT
TTACCCATAAGGATTCTCCTTTTTTTATTTTTTTGAAAAATTTACTTGCAAGACGAAGAGCT
GACAGACTTGCACCAGTCTTGAGTGTTTTTGAACCAGCTGATGAAGCTTCTTGCTCAAG
ACACGCGCATGGTCATTGAGGTCTGAAACAGATAGAGATAAATCTGCAACAGCACTGAAG
AGTGATCAATCGTAGCCACCTTGACATTGATATCATCTGCCAAGACATTGACCTTAGCC
AACAACTCATTGGTGTGATGCAAGGTCACATCCACATCTGAAGTCAAGGTTTTAATCGTC
TTTTCTGTTTCATCGATGACACGACCAAGCTTTTGTACAGTAATGATCAGATAGACCAA
AAGACAATCACAG
```

ORF Predictions:

ORF #	Start	End	Direction	Length
1	1164	1511	R	116 aa

> 3811436-3 ORF translation from 1164-1511, direction R
 VIVFLVYLIITVQKLGRVIDETEKTIKTLTSDVDVTLHHTNELLAKVNVLADDINVKVAT
 IDPLFSAVADLSLSVSDLNDHARVLSKKASSAGSKTLKTGASLSALRLASKFFKK*

Description:
 unknown

Assembly ID: 3811984
 Assembly Length: 505bp

> 3811984 Strep Assembly -- Assembly id#3811984

CTCTTGTTCAGAGAAATTTACAAAACGTTAGGAGAATAAGATGGCATTTATTGAAAAAGGT
 CAAGAAATCGATATGGAAGTCATCAAGGCTGAAACCCAATTGCTCTGCAGAAGCCTTGAGA
 CTCAAGGAAAGCCGTGACAGGGAATTGGCAGATATTATTTCAGGGGAAGATGACCGTATT
 CTCTTGGCTGATTGGTCCTTGCTCTTCTGATAATGAAGAGGCGGTCTTGGAATATGCTCG
 CCGTTTATCCGCCCTTGCAAAAGAAGGTAGCGGATAAGATTTTCATGGTCATGCGCGTGTA
 TACTGCTAAGCCTCGTACCAATGGAGACGGCTATAAAGGGTTGGTTCACCAGCCAGATAC
 TTCTAAGGCTCCAACCCTGATTAACGGCTTGCAAGGCTGTGCGCCAGTTGCACTACCGCGT
 TGATTACAGAGACTGGTTTGACAACGGCAGATGAGATGCTTTATCCGTCAAATCTGATCT
 TGGTGGATGACTTTGGTCACCTACC

ORF Predictions:

ORF #	Start	End	Direction	Length
1	134	454	F	107 aa

> 3811984-2 ORF translation from 134-454, direction F

VTGNWQILFQGKMTVFSWLIGPCSSDNEEAVLEYARRLSALQKKVADKIFMVMRVYTAKP
 RTNGDGYKGLVHQPDTSKAPTLLINGLQAVRQLHYRVDYRDWFDNGR*

Description:

PHOSPHO-2-DEHYDRO-3-DEOXYHEPTONATE ALDOLASE, TYR-SENSITIVE (EC 4.1.2.15) (PHOSP HO-2-KETO-3-DEOXYHEPTONATE ALDOLASE) (DAHP SYNTHETASE) (3-DEOXY-D-ARABINO-HEP TULOSONATE 7-PHOSPHATE SYNTHASE). - ESCHERICHIA COLI.

Assembly ID: 3857228
 Assembly Length: 1827bp

> 3857228 Strep Assembly -- Assembly id#3857228

CTCTTTTAACCGTTTTAGCGGTGACACCGAGGATTTTTTCAGGACCCAAGACTTGTCGGG
 CAACCGAAACTGGGAGTTCGTCATCTCCAATATGCAGACCAGCAGCATCAACCGCAAGAC
 AAACATCCAACCGATCATCGATTATCAAGGGGACCTGATAGGCATCTGTTATTTTCCTTGA
 CTTGTTTTGCCAGTTGATAATATTGATTGGTTGTGAGATTTTTTCTCGCAATTGGACTA
 TGGTAACCCCTGAACGGCAGGCCGTCTCAACTTTTGCAAGAAAGCTTTCCACGGAATCTT
 GATAGCGATTGGTTACCAGATATAGTCTAAGCGCTTCTCTATTCATAAACCTCTCCTTTG
 ATGGTATCTAGCCAATTTTCATCTCTTCTTAGGAGCGAAAGCTGATTGAGTACTTGGTAA
 CGAAATTCTTCCAATCCCATTCCTTGAACAACCTATTTTCTCAGCAGCGATATTGAGATAA
 GAGACTGCTAAGCAAGAACTTCAAAACCAGTCTTTCCTTGGCTGAGAAAAACAGCTGTTA

AGGCTCCAACCAAGTCTCCTGTCCCTGTTATCCAGTCTAATTCAGTACAGCCATTCTCAA
 GTACAGCAACTTGATTCTCCGAAACAATAAGGTCCTTGGGACCTGTGACTAAGAATGACA
 TACCACGATAGGTCTGACACCAGTCTTTCAAGACTTGAAGCAAATCCTCCGTTTCTTGAT
 CTTTAGCACTCGCATCGACCCCAACGCCGTGATGCTTTAATCCAACAAGACTTCGAATTT
 CTGACATGTTTCTTTAAGGACCGTAGGTCTATAGTCTAAAAGGTCTTTAACTAAGCTCT
 TACGAATGGATGAAGTCGTTACGCCAACCGCATCTACTACCATCGGGAGAGAAGATTGGT
 TTGCATACAAAGCTGCCATGCGGATTGCTTTTTCTTCTCAGCTGACAAATGCCCAAAT
 TGATGAAGAGAGCCTGGCTTTGCTTAGTAAAATCAAGAACTTCACGGGGATCATCTGCCA
 TGACAGGTTTGCATCCCAGAGCCAAAATCCCATTGCCCAGCATCTCACAAGAAATCTCAT
 TGGTCATACAGTGAATGAGGGAAGTATAGGAAAGGATTTGTCAATGCCTGCA
 TCATTCTATCCTTTTCAGCAAAGAAATATCCTTGACACTTTTTTAAAGAATTCCTGCTTGAT
 TAAAAATCTAAATGCAATAAAGGAAATCGCTGTACCAATCAAGGTTGCTCCGAAAAATCG
 AGGCGTGTAGATAAACCAACTAAGCTTAGCAGCCGATCCTGTAAAGAGCACCATAACAGG
 ATAGGAAACAATAGAACCAATAATACCTGTTCCCACAATTTCTCCCAAGGCAGAAAAGTA
 AAATTTTCGACCGTACTTATAAAAGAGACCTGCTAGAAGGGCTCCAAAAGTCGCTCCTGT
 GAGAGATAAAGGAGCTTATCGGAATACCCTTGAGTCGTCATACGGATAAAGGCTGTCACT
 GTAGCCATAGCCAAGGCATAAACAGGTCCCATCATGATTCCCGCTAGAATATTGACTACA
 CTGGACATCGGTGCCATTCCCTCAATCCGAAAGATAGGTGTAAGGACTACATCAAGGGCA
 ATCATCATAGATAAAATGGTCAATTTGTGAACTTGTAGTTGGTGCTTTCTCAAGTTTCTA
 TTCTTCTCCTTTTTCTAAAGACTGTAAATCGCTCTTCCATGTCTGGTGTTGGTAAGCCAT
 CTCCCAAACTTGGCTTCCATATGAACACTGATGTGGAAGGCATCTAGCATTTTTTGCTT
 ATCTGTCTCATCACTTTCTCGATAGAG

ORF Predictions:

ORF #	Start	End	Direction	Length
1	1141	1356	R	72 aa

> 3857228-2 ORF translation from 1141-1356, direction R
 VGTGIIGSIVSYPVMVLFTGSAAKLSWFIYTPRFFGATLIGTAISFIAFRFLIKQEFFKK
 VQGYFFAERIE*

Description:

unknown

Assembly ID: 3857842

Assembly Length: 485bp

> 3857842 Strep Assembly -- Assembly id#3857842

CTATTGCCAATCCATATAGCCTATCAGGTGGTCAATAACAACGTGTGGCCATCGCTCGTG
 GCCTATCAATGAATCCAGACATCATGCTCTTCGATGAACCAAATTCCTGCCCTTGACCCTG
 AGATGGTTGGAGAAGTAATTAACGTTATGAAGGAATTGGCTGAGCAAGGCATGACCATGA
 TTATCGTAACCCATGAGATGGGATTTGCCCGCCAGGTTGCCAACCGCGTTATCTTTACTG

CAGATGGCGAGTTCCTTGAAGACGGAACACCTGACCAAATCTTTGATAACCCACAACACC
 CTCGTCTGAAAGAGTTCCTTAGATAAGGTCTTAAACGTCTAAACTCAAACCTGCAAGGATTT
 CCTTGCAGTTTTTCTACCTCGTATTGGAATTTTTTGATTTTTTCGGAAAATTATGTTAGAAT
 TAAGTTTATGAAATGAGGTTTCCTCATACCTAGCAAGACTAGGAATAAAAATAGAAATTA
 GGTAAG

ORF Predictions:

ORF #	Start	End	Direction	Length
1	45	341	F	99 aa

> 3857842-1 ORF translation from 45-341, direction F
 VAIARGLSMNPDIMLFDEPNALSALDPEMVGEVINVMKELAEQGMTMIIVTHEMGFARQVAN
 RVIFTADGEFLEDGTPDQIFDNPQHPRLEFLDKVLNV*

Description:

GLUTAMINE TRANSPORT ATP-BINDING PROTEIN GLNQ. - BACILLUS
 STEAROTHERMOPHILUS.

Assembly ID: 3857996

Assembly Length: 1547bp

> 3857996 Strep Assembly -- Assembly id#3857996

NTCTTGGGCNCNGGGCGNNTCCTTTGAGGACNACGGTATCGATGACCTTGATCTCAAGTG
 CAAGCAGTATCTGAATCTGCAGCAGCACCTGTCCGTGCAAAAGTTCGTCCAACATACAGT
 ACAACCGCTTCAAGTTATCCAATTGGAGAATGTACATGGGGAGTAAAAACATTGGCACCT
 TGGGCTGGAGACTACTGGGGTAATGGAGCACAGTGGGCTACAAGTGCAGCAGCAGCAGGT
 TTCCGTACAGGTTCAACACCTCAAGTTGGAGCAATTGCATGTTGGAATGATGGTGGATAT
 GGTCACGTAGCGGTTGTTACAGCTGTTGAATCAACAACACGTATCCAAGTATCAGAATCA
 AATTATGCAGGTAATCGTACAATTGGAAATCACCGTGGATGGTTCAATCCAACAACAAC
 TCTGAAGGTTTTGTTACATATATTTATGCAGATTAATTTACAGAGGGACTCGAATAGAGC
 CCTCTTTTCAGGTTTTACCGTGACAATCCCTATTAAAAATTATATCAAAATCGTGAAAAT
 ATTGGAAGTATGGTAGAATGAAAATTGTCGTGTGAACGATAATACTCATTCCTTGATGA
 ATTGTGAAGCAGTTGCCCTTGGGTCGTTTTGCGAGTTGAAGTCAAGAAGAGGAAAAAAC
 AAAAAGGAGAAATACTCATCGAATTTCAATGAAACAACCTTCTTGAGGCTGGTGTACACTT
 TGGTCACCAAACCTCGTCGCTGGAATCCTAAGATGGCTAAGTACATCTTTACTGAACGTAA
 CGGAATCCACGTTATCGACTTGCAACAACTGTAAAATACGCTGACCAAGCATAACGACTT
 CATGCGTGATGCAGCAGCTAACGATGCAGTTGTATTGTTTCGTTGGTACTAAGAAACAAGC
 AGCTGATGCAGTTGCTGAAGAAGCAGTACGTTCAAGTCAATACTTCATCAACCACCGTTG
 GTTGGGTGGAACCTTACAAACTGGGGAACAATCCAAAACGTATCGCTCGTTTGAAAGA
 AATTAAACGTATGGAAGAAGATGGAACCTTTCGAAGTTCTTCCTAAGAAAGAAGTTGCACT
 TCTTAACAAACAACGTGCGCGTCTTGAAAAATTCTTGGGCGGTATCGAAGATATGCCTCG
 TATCCCAGATGTGATGTACGTAGTTGACCCACATAAAGAGCAAATCGCTGTTAAAGAAGC

TAAAAAATTGGGAATCCCAGTTGTAGCGATGGTTGACACCAATACTGATCCAGATGATAT
 CGATGTAATCATCCCAGCTAACGATGACGCTATCCGTGCTGTTAAATTGATCACAGCTAA
 ATTGGCTGACGCTATTATCGAAGGACGTCAAGGTGAGGATGCAGTAGCAGTTGAAGCAGA
 ATTTGCAGCTCCAGAACTCAAGCAGATTCAATTGAAGAAATCGTTGAAGTTGTAGAAGG
 TGACAACGCTTAATTTATACAAATAGTAATTACCTAGGAGGGCGGGGCTTAGCCCGGCTC
 TCCTATTTTCAAAAAATATAGGAGAATTAAAATGGCAGAAATTACAG

ORF Predictions:

ORF #	Start	End	Direction	Length
1	58	456	F	133 aa

> 3857996-1 ORF translation from 58-456, direction F

VQAVSESAAAPVRKVRPTYSTNASSYPIGECTWGVKTLAPWAGDYWGNGAQWATSAAAA
 GFRTGSTPQVGAIACWNDGGYGHVAVVTAVESTTRIQVSESNYAGNRTIGNHRGWFNPTT
 TSEGFVTYIYAD*

Description:

unknown

Assembly ID: 3858236

Assembly Length: 740bp

> 3858236 Strep Assembly -- Assembly id#3858236

CTATAAAAAAAGGGTAACCAGTATGGAGGATGAATGTCTGGAACCTATCTGAGAATCTCG
 GATTTTGGAAATCAGACCGATCATCATGAGATAAGGAAGGAAAGCACTTGTA AAAAGCAC
 TGTAACCACGCCAGTCCCCTGTCCCAAGAGGGTGAGGTGGTAGCGTAAAACCATGCGGAA
 AAATCCCTTTTTTAGTGGTTGAAATTCTCTCCTTGCTGCGACGTTCTTTTTTGACCTTCTC
 CTCACTATTAAGCAGGATCACGTCATAAAAACGAGGAAGGACCTTCTTTTTTGGTCAGATA
 AAGCAGGAAGAGAGTTAGTCCTATCCAAGCGAGCAGACCCAATATGGCTTCTATTGAAAA
 AGGCTCCACTGCTATTTTGTAAAAGATATGAAGAGGATAAAGGAGAAATGGAATGTCTCT
 AACTTTGTCAACAATACTTCCAAAAGTCGACTGAAGAAAGAAGATAAATATTAAAGGTAT
 GAGAACTCCTATCCCAATCATCACATTCGAAAAAATAGACTGATACTTTCTGAAGACCCT
 AGTCTGAGCCAAGAAATGTACTGCCACTACCGTCACTAAAGTAACAGAGACAAATAATAA
 GGTCAAGGACAGTAGCATCAAAGGCAAACCCAGCCAAAGAGAAGGAGCTAGACTAATATA
 GAGGGCTAGAAAAATAAGCTAGGATTGGTACAATTCCAGTTAGAGCTGGCAAGAGGACAGA
 CAGTCCTTTAGCAATTCGAT

ORF Predictions:

ORF #	Start	End	Direction	Length
1	1	261	R	87 aa

> 3858236-1 ORF translation from 1-261, direction R
VILLNSEEKVKKERRSKERISTTKKGFFRMVLRHYLTLGQGTGVVTVLFTSAFLPYLMM
IGLISKIRDSQIVPDIHPPYWLPPFL*

Description:
unknown

Assembly ID: 3858264
Assembly Length: 2219bp

> 3858264 Strep Assembly -- Assembly id#3858264
ATCGAATTTCGTTTTGCAAGTGGCGAAATGCGAACCACGTTTGTGTCTTTATAAGTTTCCA
CGTCTTCTTTGTGGACACGACCGTTTGCACCTGAGCCAGAAACGTCGTAGAGGTTTATCC
CTAAATCATCCGCTAACTTTCTAGCTGCAGGAGTCGCTCTTAGCTTGTTCATCAGCCATGA
CCTCTCCAATTCTATTTATGATACAAAGGGCGTCAAAAGCGACTGAAAAATAGGAAATCG
ACGATGGCTTCGATGAAGCCAAGGAGATTTATCTTTTTTTTCCAAGCTTTTAGCCCGTGCT
CTAATCTAAGATATTAAGGACGAAGAGCTCTGCACCTAAAAGATACAAAGTTCTCGTCAG
CTTTGTTTTATTTACATAACTTATCTTATGTAACCTCTATTCTTTGTTATAAGTTTTTCGG
ATTGCATCTTTGATACCTTCAACTGTTGGAATCATTGCACATTTTTTAGGTTTTGCGCATA
AGGCATCGGCACATCTTCTCCTGCACAACGGCGGATTGGTGCATCTAGATAGTCAAATGC
TTCTGATTCTGAAATAATAGCTGAAATTTACCGATATAGCCACTTGTTTTGTGGGCATC
GTTGACCAGAACAACCTTACCAGTCTTCTTCACTGAGTTTATGATGATATCCTTATCAAG
CGGAACAAGGGTACGTGGGTCAACAATTTCAACTGAAATTCCTTCTTCAGCTAATTCTTC
AGCAGCTTGAACCACACGGCGAAGCATTTTTTCCATAAGTGACAACCTGTTACATCCGTTCC
TTGGCGTTTTGATTTACCAACCCCAAGTGAATTTGTGTAGTCTGGATCAACTGGCACTTC
CCCTTTTTTGGTTAAATTCTGACTTGTACTCAAGTATAATAACTGGGTGTTATCACGGAT
AGAAGACTTAAGCAGGCCTTTCATGTCCGCAGGTGTTCCAGGTGCCACAACCTTAAGCCC
TGGAATGTGAGTAAACCAAGACTCTAGAGATTGTGAGTGCTGGGCGGCAGAGCCAACTCC
GTTACCAGCTGCACAACGAACAGTCATTGGAACCTGACCTTTACCACCAAACATGTAACG
TGTTTTTAGCAGCTTGTTGACGATATTGTCCATGGCAATAACAGAGAAGTCCATGAAGGT
CATATCGACGATTGGACGAAGTCCTGTCATGGCTGCTCCTGCTGCAGCTCCAGAGATGGC
AGCTTCAGAAATCGGACAGTCACGGACACGTTCTGGACCAAATTCCTCAAGCATTTCCAAC
AGAAGTACCGAAGTCTCCTCCGAAGACACCGACGTCCTTCTCCCATCAAGAACACATTTTC
ATCGCGAACGCATTTCTCCTCAGACATAGCAAGGATAATGGTGTACGGAAGGACATTGTTT
TTGTTTTCCATTTTATCTCTTTCTCCTTAGTCTGCGTAAATATCTTCAAAGGCTGATTCAA
GCGGTGGGAATGGGCTTTCCTCTGCAAATTTAACAGAAGCTTCTACTGCTTCCTTTACTT
GCGCTTGGAATTTCTTCCAATTCTTCGGCACTTGCAATGTTATTTTCAATAAGGTAATTGC
GGAGGTTTTTCGATTGGATCTTTTTTGTTCACAAATTCCTTCTTACGCGTACGATATT
TACCAGGGTCAGATGATGAGTGACCGAGCCAGCGATAAGTTACACTTTCAATCAAGACTG
GACCATTGCCACTGCGAACATGGTCTATAGCTTTCTGAAATCCTTCATAGACATCGATGA
CATTGTTACCGTCTTCGATGAACATTCCAGGAATTCATAAGCGGCGCTACGTTGATGGA
TATGTTCTATATTGGTCATTTTCTTGATATCCGCAGAGATACCGTAACCGTTGTTAATGC
AATAGAAAATGACTGGCAGGTTCCAGATAGAAGCCATGTTCACTGCTTCGTGGAAAACAC

CTTCATTGGTCGCACCATCTCCAAAGAAGCAGACAACGATTTTACCGGTATTTTGCATT
 GCTGACTGAGGGCTGCACCGACAGCGATCCCCATAACCACCACCTACGATACCATTGGCAC
 CAAGGTTCCCAGCATCAAGGTCAGCGATATGCATAGATCCACCTTTCCCTTTACAGGTTC
 CAGTGTATTTACCAAGGATTTTCAGCCATCATTCCGTTGAAGTCAATCCCTTTAGCAATAG
 CTTGCCCCGTGTCCACGGTGGTTTGAGGTAATCAGATCATCTGGATTGAGAGCTACATAG

ORF Predictions:

ORF #	Start	End	Direction	Length
1	439	1365	R	309 aa

> 3858264-1 ORF translation from 439-1365, direction R
 VTPLSLLCLRKCVRDENVFLMGEDVGVFGGDFGTSVGMLEEFGPVRDCPISEAAISGA
 AAGAAMTGLRPIVDMTFMDFSVIAMDNIVNQAAKTRYMFGGKGQVPMTVRCAAGNGVGSA
 AQHSQSLESWFTHIPGLKVVAPGTPADMKGLLKSSIRDNNPVIILEYKSEFNQKGEVFPVD
 PDYTIPLGVGEIKRQGTDTVTVTYGKMLRRVVQAAEELAEEGISVEIVDPRTLVLPLDKDI
 IINSVKKTGKVVLVNDAAHKTSYIGEISAIISESEAFDYLDAPIRRCAGEDVPMPYAQNL
 KMCNDSNS*

Description:

2-OXOISOVALERATE DEHYDROGENASE BETA SUBUNIT (EC 1.2.4.4)
 (BRANCHED- CHAIN ALPHA -KETO ACID DEHYDROGENASE COMPONENT BETA
 CHAIN (E1)) (BCKDH E1-BETA). - BACILL US SUBTILIS.

Assembly ID: 3858610

Assembly Length: 1078bp

> 3858610 Strep Assembly -- Assembly id#3858610
 CTAACCCTNGACGGGGCCGCTATCATCAGTCAAACAGCTAAAAATCTTGTCTGCAAAAGT
 CTCGATTAAGTCTTTTACAAAAGCCGATTTTCCTGGAATAACTTGGAGATTGATCAT
 CTTATCCATCAATTCAGCCGATTTCGATATTGTCTTCAGCCAGTTGCAGACTTTTACGAT
 TGATTTTGGCAATTCGTAGACATAGGTGTTGTCTCTCAAAGGAATTTTGACAATACCTAA
 CTCTTTTGATATCTCGGGATACCGTTCGCTGAGTGGCAGTGATACCTGCTTCTTTCAAATG
 TTCTACAATTTCTTCTTGCGTGCCGATTTGATAATCTGTACCAATCTTCTAATTTTTTC
 AAGTCTCTCTTTTTTATTCATTTTTAAATTGACTATGCGCCCTCTCTACTGCTTCTTTAA
 TCTCAGCAAGAATCTGATTGCTTGCTGACTTTTCTTTTTTCAAATACACTAAAAATTCAA
 TATTTCCATGTCCACCTTGGATGGGAGAAAAGTCCAAGCCAAGGACTGAAAAACCTGCCT
 CTAAGTCCATAGCTGTTACAGATTCAAGGACATTCTGATGAATCTTAGCATCTCGAATAA
 TTCCATTTTTTCCCAATCTGCTCACGTCTGCTCAAATGAGGTTTGACAAGTGCTACCA
 CCTGACCTTGATCAGCCAAGACACGGTGCAAGGCTGGCAAATCAGACTAAGGGAAATGA
 AACTCACATCAATACTGGCAAAGCTCGGCTCCTGCTCGAAATCAGTCTTTTCAGCATAGC
 GGAAATTGAACTGCTCCATGCTGACAACCTCGTGGGTCTTGGCGTAATTTCCAAGCCAAT
 GATTGGTACCAACATCGACTGCAAAGACCAACTTGGCACTATTCTGTAGCATGACATCGG

TAAACCTCCAGTAGAGGCCCCGATATCAATCGTAGTCGCGCCATCCACCGACAAATCAA
 AGACCTGCAAGGCCCTTTTCCAGTTTCAAACCACCACGGCTGACATACTTGAGTTTCTCC
 CCCTTGAGTTTAAATTCGGTGTCTCATCTGGAATTTCTCTCCTGGCTTGTCAAACCGTTC

ORF Predictions:

ORF #	Start	End	Direction	Length
1	374	949	R	192 aa

> 3858610-2 ORF translation from 374-949, direction R
 VDGATTIDIGASTGGFTDVMLQNSAKLVFAVDVGTNQLAWKLRQDPRVVSMEQFNFRYAE
 KTDFEQEPSFASIDVSFISLSLILPALHRVLADQGQVVALVKPQFEAGREQIGKNGIIRD
 AKIHQNVLESVTAMAVEAGFSVLGLDFSPIQGGHGNIEFLVYLKKEKSASNQILAEIKEA
 VERAHSQFKNE*

Description:

cytotoxin/hemolysin ORF2 tly - *Serpula hyodysenteriae*

Assembly ID: 3858716

Assembly Length: 928bp

> 3858716 Strep Assembly -- Assembly id#3858716

ACTTTCCTGACCTCTGTTTCCAAATAATCTTCCAAATGGACAGAGATCTACCGTTGTTTG
 CATCGATAGCTGAGGTCTTTTTTAGAAAATACCATCACTTTTAGAAAATATAAACACATT
 TTTCGGATAAGATTAAGGTAAAGCAGCTCGTTTATCCAGGGTCTGATGATGGTCTTCA
 CGATAAACCACATCCAATAACCAATGCATACTTTCTGCTGACCAATGACCTCGAACACTA
 TGGCAAAAGGTCATCAACATCAAGCTTAAAGTTAAAGATAAAATAGCGAACGTCTTGACT
 TGTAATACCATCTCTATCAATAGTATTACGAGTCATTCCAATTCCACGCAATTTATGCCA
 TTTGGGATGGTTTTGACACAACCACTTAACATCAGAAGACACCCAGTATTCTCGAACTTC
 AATCTATCCTCTTTCTATATTCTAACTGAAAGGACAATTCAATGATTCATTTAATAATGA
 TTAGCGCCATTGCTCTAGCCATTGGAATTGGTTACCGCACCAAAATCAATATTGGCCTGC
 TGGCTATTGCTTTTTCTTACCTCATCGCAACCACTCTCATGGGATTAAGTCCCAAAGAAC
 TTCTTCATTTTTGGCCAACCTCACTCTTTTTTACCATTTTTTAGCGTCTCTCTTTTTATA
 ACGTTGCAACAATAACGGTACTCTTGATGTTTTGGCTCAACACATTCTCTACCGCACAC
 GCACCCACCCTAACGCCCTCTACATGATTTTATACCTGATGGCAACCCTTTTGTCTGCTT
 TAGGTGCTGGATTTTTCACTACTATGGCCGTTTGCTGTCCTCTAGCGATTACCTCTGTC
 AAAAAGCGGACAAACACCCTTTGATTGGAGTCAAAGCGTCAATGGGAACCTCAGGAAGGG
 TAATTTGATAACCAAAGGAATAAAATTT

ORF Predictions:

ORF #	Start	End	Direction	Length
1	238	402	R	55 aa

> 3858716-1 ORF translation from 238-402, direction R
VSSDVKWLCCQNHKPKWHKLRGIGMTRNTIDRDGITSQDVRYFIFNFKLDVDDLPL*

Description:

unknown

Assembly ID: 3859124

Assembly Length: 847bp

> 3859124 Strep Assembly -- Assembly id#3859124

AAAAACGCACCATATCAAAAAGTTTGGATATCATGCGTCATGCTTAACTAAT
TGACTATACTTTCTATTCAAATGAGCTTTTAACCAATTGATTGAGCCAATCCACTCTTAA
AACCAAAGGAGCAATTTCTCGGCTTAGCTGACTCTTCTCGGAATCTGAACCATGTACAAC
ATTTTGGATAATCTCATTTTCTCCAGCAGCTTTTGCAAAATCACCTCGAATAGTGCCTGG
TAAAGCTTCTTCTGGACGAGTTGCACCCATCATGGTCCGCCAAGTTTTCGATTACTTTGGG
ACCAGAAATGACACCCACAAGAACTGGACCTGAAGTCATGAATTCACGAATCGGTGGGTA
AAAACCTCTGACCAACCAAGTCCTGATAGTGCTGGTCAATCAACTCTTCTGAAAACCTGTG
AACGAACTCCAATTTTTCGATTGTAAATCCACGTTGTTTCGATGCGCTTTAACACTTCAC
CCACTAGCCCTCTTTTACACCATCTGGTTTGGATGATAAAGAATGTTTGTTCATACCCG
TCTCCTTTGTGAGCTTCTTTCTTTTATTTTACCACATCTCGTGGAAAAATGGAGAAAGTT
TTCAGAAGAGAGAATGAGAGAACCCTCGGGTTCTCTCATTTCTCTCTTATTCTACTGTTTC
TTCCACAGTGTCAACGGCAGTATCCACAACACTACTTCTGTTGTTTCTTCATTTCTTCTTC
CTCTACTGGAGGATTAAGGTATTCTTCTTTCGTTGACAGCATGTGGTTCAAGGTTACGGTA
ACGGGCCATACCAGTACCAGCTGGGATGATCTTACCGATGAATAACATTTTCTTTAAAT
TCCAAGG

ORF Predictions:

ORF #	Start	End	Direction	Length
1	73	453	R	127 aa

> 3859124-1 ORF translation from 73-453, direction R
VDLQSKNWSFVHRFSEELIDQHYQDLVGQSFYPPPIREFMTSGPVLVGVISGPKVIETWRT
MMGATRPEEALPGTIRGDFAKAAGENEIIQNVVHGSDSEKSQLSREIAPLVLRVDWLNQL
VKSSFEE*

Description:

NUCLEOSIDE DIPHOSPHATE KINASE (EC 2.7.4.6) (NDK) (NDP KINASE)
(ABNORMAL WING DI SCS PROTEIN) (KILLER-OF-PRUNE PROTEIN). -
DROSOPHILA MELANOGASTER (FRUIT FLY) .

Assembly ID: 3859244

Assembly Length: 578bp

> 3859244 Strep Assembly -- Assembly id#3859244

ACAACCTAACTACCGNCTAATTCAGCGCGAACTTCTGCAGTAGCTGCTTCAACAACCTTCA
CGACGTGAAAGGATGAAGCGGTTTTCTTTAGCGTTAACTTCTTTGATTTTAGTATCAAAT
TCTTGACCTACAAAACGCTCAGCGTTACGTACGAAACGAGTATCCAACATTGAAGCTGGG
ATAAATCCACGAACACCTTCAAATTCTACTGAAAGTCCACCTTTAACGGCAGCGTTTCTT
TTAACAGTAACAACCTTCTTCTTCGCGACCAACAAGTTTGTCCCATGCTTTGCGAGCTTCA
AGGCGTTTTTTTAGATGACAAGGTATGTAAGTGTATCAGTATCTTTACCAACTACTTGACG
AAGTACAAGAACATCCAATACTTCTCCTACTTTAACAAAGTCATTGATATCTGCATCACG
ATCGTTTTGTCAATTCGCGAAGAGTCAAGACACCTTCAACACCAGTTCCCAGAAGAATGC
AACGTTAGCTTGAGTCGCATCAACTGTCAATACTTCAGCACTAACACATCACCAGTCTCA
ACTTGACTNACGCTATTGAGCANATCTTCAAATTTCGAT

ORF Predictions:

ORF #	Start	End	Direction	Length
-----	-----	-----	-----	-----
1	310	462	R	51 aa

> 3859244-2 ORF translation from 310-462, direction R

VLKGVLTRELTLNDRDADINDFVKVGEVLVDLVLVRQVVGKDTDTVTYLVLI*

Description:

unknown

Assembly ID: 3859250

Assembly Length: 888bp

> 3859250 Strep Assembly -- Assembly id#3859250

GTAGTTATAGTAGGGGTCGGATTGAAATGCCACNGCGCTTCTTGAGATTCTGATACCGT
TTAAAATAGCGTTGGGCATTCTGGTTGGGAGTCAGAGCCTTATCAAGCGCAATCATGATA
GGTTGGTTGGTATAGTAGTTGTCTAGGATAACCTGGTTCTTGGTCGTTAGGCACCTGGTG
GAGGAAGGTTGTCAGCAATTCTCCTTTTTTGACGAAATTCCTCAGCGTTGTCTGTGCCAG
TAACTATTTTTCTGTTTTTTTGAGTTTGTGTGCGTTTTTCTGAAGTTCATTTTCAACACG
ACGAATCAGTTCACTGGCCTGCTGTTTTGACGCGGTCGCGCTCAGCCTTATCCTTATAGTA
GGTGTCCAACAAATCAGAAAGATTTGCAAAAGGCTCTCCACCTGATTTGCAAAAGGAAC
TGGACTGAAGGAAGTCTCAGTCAAGCATGGCTTGGTTTCCTGATTGAAAAAATTTTCGGAA
AGCGGAAAGTTTTTCACTAACCAGTATCCTTTCCAATTCATTGCCGTATCGCGTCCCAG
ACCTTGAAAGAGGCTTTGAAGATTTTTTTGCTGTTAGTTCTTGGGTTTGCAGGATTTCAAA
GAGCTTTTCATCCTTGATAGTAAAAGGATTGAGAGATTCTGTACTTGGCGGAGCGATATA
GGTCGATCCTGGAAGTAAGGTGCGGTAGCTATTTTGTGAAAAGCCGACGTGTTTGATAAC
TTCGAGGATTTTATGACTGCTTTTATCCGACCAGTTAGAATATTACTGTGTTTCCCCATA
ATTTGATAATCAAGGTAGCCTGGATATGGTCTCCAATCTCGTTTTTTATTGGAACTGTA

ATTTCCACAATACGGTCATTTTCCACTTGCTCAATCGACTCAATCAGG

ORF Predictions:

ORF #	Start	End	Direction	Length
1	244	402	R	53 aa

> 3859250-1 ORF translation from 244-402, direction R
VGEPPFANLSDLLDTYYKDKAERDRVKQQASELIRRVENELQKNRHKLKKQEK*

Description:

STRFBP5A NCBI gi: 496253 - Streptococcus pyogenes.
Fibrinogen/Fibronectin binding protein

Assembly ID: 3859588
Assembly Length: 513bp

> 3859588 Strep Assembly -- Assembly id#3859588
ATCGAATTTTGTTCCTTTCATAGAGAGCTACCTGAGTTCTATTCAAGCTCAGGTAGTACTT
TCTTATAAACTAGACAACTAACTGTCATTCTACCATCAGATTACAAGACATCATCGTCA
CTCACCTTGGGAATTCAATGTCGTACCCCAATGGGTAATTTTACGGTGGGGTTGAGCTAAA
ATTGGTCTGTTTTTCATAGATTGTTTGCCATCTATTCCATAGTAGGCCCGTCTTTTTCTCA
ATCTTAACTCGCAGATTTCTCATATTTTCTTTGATTGGGAGGTTGAGGACAAAACCTGCA
GTCTGGTTGCGACCGTTTCCTTCCCAAGAATGACTACGAACAACCTGGTTTCCATCTTTA
TCTACTGGAACCTTCTTCCCAAGTTATGGAGTAGCGGGCAATGTAAGCTCCACTGTGTTGA
ATTATCAATGTTTTATCTTTTACAGGGAGTCTGACTGATTGGTTGAACTGGCTTAGAAAC
TTGTGTCGCCGTTTCAGCATTCGTAGCTATAAA

ORF Predictions:

ORF #	Start	End	Direction	Length
1	102	443	R	114 aa

> 3859588-1 ORF translation from 102-443, direction R
VKDKTLIIQHSGAYIARYSITWEEVPVDKDGNOVVRSHSWEGNGRNQTAGFVLNLPIKEN
MRNLRVKIEKKTGLLWNRWQTIYENRPILAQPHRKITHWGTTLSKVSDDDVL*

Description:

PNEUMOLYSIN (THIOL-ACTIVATED CYTOLYSIN). - STREPTOCOCCUS
PNEUMONIAE.

Assembly ID: 3859774
Assembly Length: 214bp

> 3859774 Strep Assembly -- Assembly id#3859774
 ATCGAATTCTAACATGTGCTTCTCCTTCTATTGTTCCCTATCTTTAAAATCTACTCCTTCA
 TGCTCCAAGAGCCAAGCTTTCTTTTCCACTCCTGCAGCATAACCTGTCAGACGCTTGCCT
 GCTCCCAACACACGATGACAAGGTACTAGGATAGACCAAGGATTGCGTCCCACTGCTCCA
 CCAATTGCTTGAGCAGAAGCCACTTGCAGGTCTT

ORF Predictions:

ORF #	Start	End	Direction	Length
1	9	131	R	41 aa

> 3859774-1 ORF translation from 9-131, direction R
 VLGAGKRLTGAAAGVEKKAWLLEHEGVDFKDRNNRRRSTC*

Description:

GLUTAMATE RACEMASE (EC 5.1.1.3). - ESCHERICHIA COLI.

Assembly ID: 3860140

Assembly Length: 1084bp

> 3860140 Strep Assembly -- Assembly id#3860140
 CTCCAGCAATGGATCCAAGTATGATGGGCGGGATGATGTAAGCTTTCTATAGAAAACACC
 TTATAAAAAACACGAAAGGAGGGAATGACTAACCCTTCTTTTTTATAATATTCACTTCTAA
 GATTGATGGTGAGCTCTCCTAAGTATATGATAAAATAAGACTAGAGGAAAGGAGAAGAA
 CATGATCGATGTACAAGAAATCTGTGCAAGATGACCCCCAATCAGAAGATTAATTATGA
 CCGTGTCATGCAGAAAATGGTACAAGCATGGGAAAAAATGAGTAGCGGCCAACCATCT
 CGTGTCATGTTTGCTGTGCCCCCTTGTAGTACCTATACACTAGAATATTTGACCAAGTATGC
 AGATGTGACCATCTATTTTGCCAATTCTAATATCCATCCCAAGGCAGAATACCATAAGCG
 GGTCTATGTACCAAGAAATTTGTTAGTGATTTTAATGAGCAGACAGGAAATACGGTTCA
 GTACCTAGAAGCTCCCTACGAACCCAATTAATACCGAAAACCTAGTTAGGGGGCTAGAGGA
 GGAGCCCGAAGGTGGCGACCGTTGCAAGGTTTGTTTTGACTACCGACTGGATAAAACAGC
 GCAAGTGGCTATGGACTTGGGCTTTGACTACTTTGGTTTCAGCCTTGACCATCAGTCCTCA
 TAAGAATTCTCAAACCTATCAATAGCATCGGAATCGATGTGCAAAAAATTTACACGCCCCA
 CTATCTTCCCAACGATTTCAAGAAAAATCAAGGCTACAAACGTTTCAGTAGAGATGCGTGA
 GGAGTATGATATCTATCGTCAATGTTATTGTGGCTGCGTCTATGCAGCCCAAGCCCAGAA
 TATTGACCTGGTTTAAGTTGAGTAGGACGCCACAGCATGCTTGCTGGATAAGGATGTTGA
 GAAAGACTATTCTCATATCACATTTATAGTAGATTGAAACTAGAATAGTACACCTTTACT
 TCTCAAACATTGTTAGAAATCGATTTCGGCTGTCCCTTATTTTCAATTTTAATACTGGTACG
 AAATTAGATATATCAATGATAACTTGCCTCAAGGTAGGTTTTTTGATAGTAGAAAAGCGA
 TAGA

ORF Predictions:

ORF #	Start	End	Direction	Length
1	302	511	F	70 aa
2	605	856	F	84 aa

> 3860140-1 ORF translation from 302-511, direction F
VHVCCAPCSTYITLEYLT KYADV TIYFANSNIHPKAEYHKRVYVTKKFV SDFNEQTGNTVQ
YLEAPYEPN*

Description:
unknown

> 3860140-2 ORF translation from 605-856, direction F
VAMDLGFDYFGSALTISP HKNSQTINSIGIDVQKIYTPHYLPNDFKKNQGYKRSVEMREE
YDIYRQCYCGCVYAAQAQNIDL V*

Description:
unknown

Assembly ID: 3860206
Assembly Length: 1124bp

> 3860206 Strep Assembly -- Assembly id#3860206
ATCGAATTCATTGACTGCCTGAAAAGACTTCAACTCGTCTGCCTGATAACCGAAAGACTT
GGTTACTTTTGATACCTGATACGGACTCCTGTACCTTGTTATTGAGTTCAGAAAAAGCAGC
TTGGGATTTCGCCAAAGGCCTTATGAGTCTTTCTCCCTAGGCGACTAGTCGTATAGGCCAT
GAAAGGTAGGGGGAGAATGGCAACAAGAGTCATCTGCCATGAGATGCTAAAGAGCATGGT
CAACAAAGTCACCAGAGCCGTGATAGAGGCATCCACCGCAGACATGACACCGCCACCTGC
TAAACGAGTCAAGGAATTGATATCATTGGTTGCGTGTGCCATCAGATCACCCGTCCGATA
GGTTTGATAAAAGGCTGACGACATTTTTGTGAAATGCTTAAACAAGCGAGACCGCATGAT
CTGTCCCAAGCAATAAGAGGTCCCAAGGATATACATACGCCACACATAGCGCAAATAGTA
CATACCAAAGGCTGCAAGTAGCAAGTAAATAGGCTAAGAAGGAGGTCCTGCTGGGTAA
TTGCCCCGATGTGATGGCATCAATAACCCGCCCCATAACCATAGGAGGAATGAGATTGAG
GACGGAAACCAAGACCAGGGCCACAATCCCGACTAGATAACGGCGTTTTTCTA ACTTGAA
AAACCACCAAAATTTTTGAATAATGGACATAAAATCCCTTTCTGGATTGCAAATAGAAAC
CTGAGGCCAATACTCAATGGAAAATCAAAGAGCAAAC TAGGAACTAGCCGCAGGCTGCT
CAAAGCACTGCTTTGAGGTTGTAGATAGA ACTGACGAAGTCAGTAACCTACATACGGCAA
GGCGACGTTGACGCCGTTTGAAGAAATTTCCGAAGAATACAAGACCCCAGGTTTTTCTTA
TTTATAAGTTACCACTGTAACAGCACCCCTTGTCATATTCAGCAATAAAGATATTGGCTAC
ATTGTCATGCCCTTGTTTACTGAGGTTATCAAGCAACCACTCCTCGCTACGAACAATCGA
TCCCAAGACATCTACTTGAATCACACCGTCAGTCACA ACTGGATACTTAGGATTTTCATC
TCCCATTTGCACAACGATGAGTTGCCCATTTTGCTCTTGACACAG

ORF Predictions:

ORF #	Start	End	Direction	Length
1	898	1056	R	53 aa

> 3860206-2 ORF translation from 898-1056, direction R
 VTDGVIQVDVLGSIVRSEEWLLDNLSKQGHNDVANIFIAEYDKGAVTVVITYK*

Description:

unknown

Assembly ID: 3860270

Assembly Length: 1242bp

> 3860270 Strep Assembly -- Assembly id#3860270

TTACCTTCATTGCAGCCATTATTGGTCTTGTGTCAGCCAGATTTTAAGTATTCTTTATA
 AGACACCTGCTGTGGTCTTTATCTTGGCCATTTTGGCACCGCTGGTTCAGGTTATCTCT
 CCTACCGAACAACCTGCCTTTTTTGTGACAGGGGACTATAATAAAGCACTGGCAAGTGCGA
 CCTTGGTTGTCATGTTGGCTTTGGTAATCTCTATTGGAATGGCTAGCGGAACAGTGATTC
 TCAGACTGTATCATTATATAAAAACACATCGAGTATCGTAGACTTTACAGAAATAAAGA
 ATTTTCTGAAAAATGAGATAAAATAAATTAACAACGCTTTCTATATGTGCGAGAATACCGC
 ACTTATGAAGAAATTGCGGCTGATTTTGGTATCCACGAAAGCAACTTAATCCGTCGGAGC
 CAATGGGTTGAAGTAACTCTTGTTCAAAGTGGTGTTACGATTTCAAAAACCTCATCTTAGT
 GCTGAGAATACGGTGATTGTGGATGCAACAGAGGTAAAAATCAATCGCCCTAAAAAACAA
 TTAGCGAATGATTCTGGTAAAAAGAAATTTACAGCTATGAAGGCTCAGGCGATTGTCACA
 AGTCAAGGGAGAATTGTTTCTTTGGATATCGCTGTGAACCTATTGTCATGATATGAAGTTG
 TTCAAAATGAGTCGCAGAAATATCGGACAAGCTGGAAAAATCTTGGCTGATAGTGGTTAT
 CAAGGGCCCATGAAGATATATCCTCAAGCACAACTCCACGTAAATCCAGCAAACCTCAAG
 CCGCTAATAGCTGAAGATAAAGCTTATAACCATGCGCTATCCAAGGAGAGAAGCAAGGTT
 GAGAACATCTTTGCCAAAGTAAAAACGTTTAAATGTTTTCAACAACCTATCGAAATCAT
 CGTAAACGCTTCGGATTACGAATGAATTTGATTGCTGGCATTATCAATTATGAACTAGGA
 TTCTAGTTTTGTCAGGAAGTCTATTATTTTCCTTATTGTCTGTAAGTCTACTGACCTTGTT
 GTTTATCCCAGTCATGGTTTCTAGTTCGGGCTCAGAGTTTCAAAGTGGATGGCAAGAGCA
 TCAATTGATTGCTGAGAAGGTAGTAAACACTTGACAAGACATTTGATAAGGATGTCAG
 AAAAATTCGACCAGTCAGTTTTATCAAAAATTTGTAGATGAGATGGGAAGGATTTACTC
 AGGAAATTTGATCCTCCCAGGAGCTGATAACTGTGAATGGAG

ORF Predictions:

ORF #	Start	End	Direction	Length
1	346	966	F	207 aa

> 3860270-1 ORF translation from 346-966, direction F

VREYRTYEEIAADFGIHESNLIRRSQWVEVTLVQSGVTISKTHLSAENTVIVDATEVVKIN
 RPKKQLANDSGKKKFHAMKAQAIVTSQGRIVSLDIADVNYCHDMKLFKMSRRNIGQAGKIL
 ADSGYQGPMKIYPQAQTPRKSSKLKPLIAEDKAYNHALSKERSKVENIFAKVKTFKMFST
 TYRNRKRFRGLRMNLIAGIINYELGF*

Description:

ISL2 protein - *Lactobacillus helveticus* (Probable transposase)

Assembly ID: 3860438

Assembly Length: 1575bp

> 3860438 Strep Assembly -- Assembly id#3860438

GTGATGGGGCCTCAGGGAAATGGTTTTGGCTTGTCTGACCTTGATGAGCAGAATCAGGTT
 CTCCTTGTGGTGGTGGGATTGGTGTTCACCCCTTGCTTGAGGTGGCCAAGGAATTGCAT
 GAACGTGGAGTGAAAGTAGTGACAGTCCTCGGTTTTGCTAATAAGGATGCTGTTATTTTG
 AAAACGGAATTGGCTCAGTATGGTCAGGTCTTTGTAACGACAGATGATGGTTCTTATGGC
 ATCAAGGGAAATGTTCCGTTGTTATCAATGATTTAGATAGTCAGTTTGATGCTGTTTACT
 CGTGTGGGGCTCCAGGAATGATGAAGTATATCAATCAAACCTTTGATGATCACCCAAGAG
 CCTATTTATCTCTGGAATCTCGTATGGCTTGTGGGATGGGAGCTTGCTATGCCTGTGTTC
 TAAAAGTACCAGAAAGCGAGACGGTCAGCCAACGCGTCTGTGAAGATGGTCCTGTTTTCC
 GCACAGGAACAGTTGTATTATAAGGAGAAAATTATGACTACAAATCGATTACAAGTGTCT
 CTACCTGGTTTGGATTGAAAAATCCGATTATTCCAGCATCAGGCTGTTTTGGCTTTGGA
 CAAGAGTATGCCAAGTACTATGATTTAGACCTTTTAGGTTCTATTATGATCAAGGCGACA
 ACCCTTGAACCACGTTTTGGGAATCCAACCTCAAGAGTGGCAGAGACGCCTGCTGGTATG
 CTCAATGCAATTGGCTTGCAAAATCCTGGTTTAGAGGTTGTTTTGGCTGAAAAGCTACCT
 TGGCTGGAAAGAGAATATCCAAATCTTCCTATTATTGCCAATGTAGCTGGTTTTTCAAAA
 CAAGAGTATGCAGCTGTTTCTCATGGGATTTCCAAGGCAACTAATATAAAAGCTATCGAG
 CTCAATATTTCTTGTCCCAATGTTGACCACTGTAATCATGGACTTTTGATTGGTCAAGAT
 CCAGATTTGGCTTATGATGTGGTGAAAGCAGCTGTGGAAGCCTCAGAAGTGCCAGTTTAT
 GTCAAATTAACCCCGAGTGTGACCGATATCGTTACTGTCGCAAAAGCTGCAGAAGATGCG
 GGAGCAAGTGGCTTGACTATGATCATACTCTGGTGGGATGCGCTTTGACCTCAAAACCAG
 AAAACCAATCTTGGCCAATGGAACAGGTGGAATGTCAGGTCCAGCAGTTTTCAGTAGCC
 CTCAAACCTCATCCGCCAAGTAGCCCAACAACAGACCTGCCTATCATTTGGAATGGGGGGA
 GTGGATTTCGGCTGAAGCTGCCCTAGAAATGTATCTGGCTGGGGCATCTGCTATCGGAGTT
 GGAACAGCTAACTTTACCAATCCTTATGCCTGCCCTGACATCATCGAAAATTTACCAAAA
 GTCATGGATAAATACGGTATTAGCAGTCTGGAAGAACTCCGTCAGGAAGTAAAAGAGTCT
 CTGAGGTAACTGCAATCAATCTGTTCTTGATTTTTTATTAGTTTGTAATATGAATTTAG
 GAGAATTTTGGTACAATAAAATAAATAAGAACAGAGGAAGAAGGTTAATGAAGAAAGTAA
 GATTTATTTTTTTAG

ORF Predictions:

ORF #	Start	End	Direction	Length
-----	-----	-----	-----	-----

1	1	276	F	92 aa
2	460	1128	F	223 aa

> 3860438-1 ORF translation from 1-276, direction F
 VMGPQGNQFDLSDLDEQNQVLLVGGGIGVPPLLEVAKELHERGVKVVTVLGFANKDAVIL
 KTELAQYGVFVTTDDGSYGIGKGNVPLLSMI*

Description:
 unknown

> 3860438-3 ORF translation from 460-1128, direction F
 VKMVLFSAQEQLYYKEKIMTTNRLQVSLPGLDLKNPIIPASGCFGFGQEYAKYYDLDLLG
 SIMIKATTLEPRFGNPTPRVAETPAGMLNAIGLQNPGLLEVLAEKLPWLEREYPNLPPIA
 NVAGFSKQYAAVSHGISKATNIKAIELNISCNVDHCNHGLLIGQDPDLAYDVVKAHAVE
 ASEVPVYVKLTPSVTDIVTVAKAAEDAGASGLTMIILWWDAL*

Description:
 DIHYDROOROTATE DEHYDROGENASE (EC 1.3.3.1) (DIHYDROOROTATE
 OXIDASE) (DHODEHASE). - BACILLUS SUBTILIS.

Assembly ID: 3860544
 Assembly Length: 776bp

> 3860544 Strep Assembly -- Assembly id#3860544
 CTAAGATATCAGAATAACAACGAAATCGAAGCATTAACAAACAAATATTACTTCTAAGAAT
 AGCGAGATTGATAGTCAACAAAGCAATATTAAGGATATGACCGTACCTATAATGATCCAA
 CTTCTCAGGCTTATAATATTTATGCTCAATTAATTAGTGAGTTAGGTACTGCTCGTTCAA
 ACAACAATAAAAGTATTACAGAGCTTGAGGCTAATCTTGGAGTGGCAACAGGTCAAGATA
 AAGCTCATAGTATATTAGCGTCAAATGAAGGTACTCTGCATTATCTGGTACCTTTGAAAC
 AAGGAATGTCTATTTCAGCAGGGGCAAACGATAGCAGAAGTTTCAGGGAAAGAAAAAGGTT
 ACTATGTAGAGGCTTTTGTACTTGCGAGTGATATTTCTCGTGTTTCAAAGGAGCAAAAG
 TTGATGTTGCTATTACTGGTGTGAATAGTCAAAAATATGGAACACTAAAGGGACAAGTCA
 GACAGATTGATTTCAGGAACAATTTCCCAAGAAACGAAAGAGGGGAATATTAGCCTCTATA
 AAGTCATGATAGAATTAGAAACCTTAACCTCTAAAACATGGAAGCGAGACGGTCATACTCC
 AAAAGGATATGCCAGTTGAAGTGCGGATTGTCTATGATAAAGAAACCTATCTTGATTGGA
 TTTTAGAAATGTTAAGTTTCAAGCAATAATTGGTTTTAAACCTTAGGTAACCTATAAAAA
 CAAATAAGGTAGAGAAAGGATATTTTATCTAAGTTAGCTCACATTACTGCCATTCC

ORF Predictions:

ORF #	Start	End	Direction	Length
1	222	689	F	156 aa

> 3860544-1 ORF translation from 222-689, direction F
 VATGQDKAHSILASNEGLHYLVPLKQGMSIQQGQTIAEVSGKEKGYVVEAFVLASDISR
 VSKGAKVDVAITGVNSQKYGTLKGQVRQIDSGTISQETKEGNISLYKVMIELETTLTKHG
 SETVILQKDMPEVRIVYDKETYLDWILEMLSFQK*

Description:

unknown

Assembly ID: 3860558

Assembly Length: 1487bp

> 3860558 Strep Assembly -- Assembly id#3860558

CTGGCCTTTCTCCACCAAATTTGTTTCCTTGAGGGAAGGAAGTCAGAACACTAGCCGTTGC
 ATCTTCCTTTTGCTTTTCAATCGTAATTCCAGATAATTTTCCCATTCTTTTGGTGACC
 CCGGGAGGCAGGATTGAATGGCTTGAGGGAAATGACAACTTGTCCTAGCAAGAATGGTC
 AAGGCACCTCCGTCTACAATCAAAATCTGATTTGGGCTTAAATTAACAAAGACCTGTTTT
 ACTAGATTTTCTCCAGAAGCATCGTCTCGTAAACCAGGCCCCAGCAAGATAACTTCTGCC
 TTCTCCAATTGCTCTTTTAACAATTGCTGGTCTTGAAGAGAAAAGGCCATAGGCTCAGGT
 AAATGGCTGTGCAGAGCCGGGATATTTTCCCTGTCCGTCCAACGGTCACCAATCCTGCA
 CCGCTTTTTTACAGCTGCTAAAGCAGCCATGATGATGGCACCTCCATAAGGATAAGTACCA
 CCAAGCAGCAGCAGACGACCATAATCTCCTTTATGACTTGAACGAGAACGTTCAATAATA
 ACTTTTTCTAGTAAGGTTTGATTAATCACTTTCATCCTTTTTTCCCTCTCACTTTTATTAT
 ACAACAAAAAGGAGACGCAGACCTCCTTTTGTAATCTTATATCTAAAATTTAATATTCAT
 TTCTGCCATTTTAGATATAGCTATAGAAAATACACTCTATTAATCGAATGTTTCTCTTAT
 TTTCTATCCAATGTCCGAAGTGCTGCTTGATAAGTTTGCTCCATCAGCATGGTAATGGTC
 ATAGGACCGACACCTCCAGGGACTGGCGTGATATGGCTAGCAAGTGGTGCAACTGCCTCA
 TAATCAACATCTCCACAGAGCTTCCCATTTTCATCTCGGTTTCATCCCAACGTCAATGACA
 ACCGCACCTGGTTTGACAAAGTCAGCAGTCACAACTTGGCGCGGCCGATTGCGACTACA
 AGAATATCTGCTTTAGCAGCCACCTTGGCAAGATTATGAGTTCGTGAGTGGGCCAAGGTT
 ACTGTGCGATTTTGTAGCCAAAAGAAGCTGAGCCATAGGTTTTCCAACGATATTTGAACGA
 CCGATTACGACCGCATTTTACCTTCCAAGTCAATCCCATATTCATGAAACATTTCCATA
 ATTCCTGCAGGTGTCGAGGGAATCATGACTGGATGTCCAGACCAAGACGTCCCATGTTT
 AGGGGATGGAAACCATCCACATCCTTTTCTGGGTCAATGGCTAATAAAACCGCCTCTTCA
 TCGATATGTTTTGGTAATGGCAACTGGACCAAAATCCCATGCCAAGCTGGATCCTGATTA
 TATTTAGCAATCAGGTCTAACAATTCCTCTTGAGTAATGGTCTCTGGAACCTCGCACTACT
 TCGGTACGGGAACCAGCCGCAAGAGCTGACCTCTCCTTGTTGCGAACGTTAAACTTGGCT
 GGCTGGATTATCCCCAACCAAAATCACTACCAAAACCAGGCACTAGAG

ORF Predictions:

ORF #	Start	End	Direction	Length
1	717	1376	R	220 aa

> 3860558-2 ORF translation from 717-1376, direction R
VRVPETITQEELLDLIAKYNQDPAWHGILVQLPLPKHIDEEAVLLAIDPEKDVDGFHPLN
MGRLWSGHPVMIPSTPAGIMEMFHEYGIDLEGKNAVVIIGRSNIVGKPMQQLLLAKNATVT
LAHSRTHNLAKVAAKADILVVAIGRAKFVTADFKPGAVVIDVGMNRDENGKLCGDVDYE
AVAPLASHITPVPGGVGPMTITMLMEQTYQAALRTLDRK*

Description:

5,10-methylene-tetrahydrofolate dehydrogenase (fold) homolog -
Haemophilus infl uenzae (strain Rd KW20)

Assembly ID: 3860568

Assembly Length: 1634bp

> 3860568 Strep Assembly -- Assembly id#3860568

CGTGCCTTGCCCAATGATCCAAAAATCTTGATTTTCAGACGAGTCGCTTCAAATTTTCGGCC
CCTGGACCCTTAAGACCAACCCAAGCAGATTTTGGCCCTTGGTTGCAAGATTTGAACCAA
AAATTAGGCTTGACTGTTGTCCTGATTACGCATGAAATGCAGATTGTCAAAGACATTGCC
AACCGTGTGTCAGTTATGCAGGATGGGCATTTGATTGAAGAGAGTAGTGTGCTTGAAATC
TTCTCAGACCCTAAACAACCTTTGACTCAAGACTTTATCTCAACAGCTACAGGTATTGAC
GAAGCCATGGTCAAAATCGAGAAGCAAGAAATCGTGGAACACTTGTCTGAAAACAGTCTC
TTGGTGCAACTCAAGTACGCTGGATCTTCAACAGACGAGCCACTTTTGAATGAATTGTAC
AAGCATTATCAAGTAATGGCTAATATTTCTCTATGGGAATATCGAAATCCTCGATGGTACT
CCTGTTGGAGAATTGGTGGTGGTCTTGTTCAGGTGAAAAAGCAGCGCTGGCAGGTGCTCAA
GAAGCCATTCGTCAAGCAGGCGTACAGTTAAAAGTATTGAAGGGAGGACAGTAAGATGGA
ATCATTTGATTCAAACCTATTTACCAAATGTCTATAAGATGGGTGGTCTGGTCAGGCAGG
CTGGGGAACAGCTATCTACCTAACCCTCTATATGACAGTTCTTTTCCTTCATTATCGGAGG
CTTCTTGGGGCTAGTGGCAGGTCTCTTTCTCGTCTTGACAGCGCCAGGTGGTGTCTTGGA
GAATAAAGTCGTATTCTGGATTTTAGACAAAATTACCTCAATTTTTCGTGCGGTTCCCTT
TATCATCCTCTTGGCAATCTTGTTCACCACTTTCTCACTTGATTGAAAAACAAGTATCGG
GCCAAATGCAAGCCCTTGTCCCACTTTCTTTTGCAGTCTTTGCCTTCTTTGCCCGTCAGG
TGCAGGTTGTCTTGGCTGAAATGGATGGCGGTGTCATTGAGGCGGGCTCAAAGCGAGCGG
AGCGACTTTCTGGGACATCGTGGGTGTTTACCTATCAGAAGGTCTTCAGATTTGATCCG
TGTGACGACTGTGACCTTGATTTCCCTTGTTGGGGAAACAGCTATGGCCGGTGCGGTTGG
AGCTGGTGGTATCGGTAACGTAGCCATCGCTTATGGATTTAACCGCTACAATCACGATGT
GACCATCTTGGCAACCATCGTTATCATTTTGATTATCTTTGCAATCCAATTCTTAGGAGA
TTTCTTGACTAAGAAATTGAGCCATAAATAAAAAAGAGCCGTGTGGCTCTTTTAACTGA
TCAGATTTTCTGGGCAAATTTTCTCAAGGCTTGTTCCAATCAAGGCACCCACTAGGGC
TCCGATGACAATACTTGGGATAAATAGAAGGACAGTTCCAGGGTTTGGAGCGACCATGAT
GCGGTGATATATTCTTGGGATTTTCCTCTTGCCAGAAGAGTAGCCATATAGGCTTTGGG
CGCAATCCACATAAGCAAGATTGGTCCTGTTGTACTAAAGCGAAAATAATGAAAGAAAG
GAAGTTCTTTGTTTTGTCCTTGATTTTTCCTAAATGAGCTACTCCATCTGCTAGGAGGCC
ACAGATAATTTCGAT

ORF Predictions:

ORF #	Start	End	Direction	Length
1	1040	1291	F	84 aa

> 3860568-3 ORF translation from 1040-1291, direction F
 VGVYLSEGLPDLIRVTTVTLLISLVGETAMAGAVGAGGIGNVAIAYGFNRYNHDTVILATI
 VIILIIFAIQFLGDFLTKKLSHK*

Description:

unknown

Assembly ID: 3860582

Assembly Length: 1087bp

> 3860582 Strep Assembly -- Assembly id#3860582

GGAATCATGATGATGTCACCTGCTAAATGGTTTCTTAGAAAAATATTCCTGAGCGCTTA
 CAGATTAGTTTGGGCTTGCTGATTTTATCATTGAGCGGTACAGCTCCCTTCTGGTACCAA
 GCCTATCCCTTTGTCTTTGGAACACGGCTTCTCTTTGGTTTGGGTCTTGGGATGATCAAT
 GCCAAGGCCATTTCTATTATCAGTGAACGCTACCAAGGAAAAAGGCGAATTCAGATGTTA
 GGGCTACGCGCTTCTGCAGAGGTCGTTGGAGCTTCTCTCATTACCTTGGCCGTCGGTCAA
 GTTGTGGCCTTTGGTTGGACAGCTATCTTTCTAGCCTATAGTGCTGGATTTTGGTGCT
 GCCCTTTATCTGCTCTTTGTCCCTTATGGAAAATCAAAGAAAGAAGTCAAGAAAAGAGC
 GAAGGAAGCAAGTCGTTTAACTCGAGAAATGAAAGGCTTGATTTTACCTTAGCTATCGA
 AGCGGCAGTTGTAGTTTGTACCAATACAGCTATTACCATCCGTATTCCAAGTTTGATGGT
 GGAAAGAGGATTGGGGGATGCCAGTTATCTAGTTTGTCTTAGTATCATGCAGTTGAT
 CGGGATTGTGGCTGGGGTGAGTTTTCTTTCTTGATTTCTATCTTTAAAGAGAACTGCT
 CCTCTGGTCTGGTATTACCTTTGGCTTGGGGCAAATCGTGATTGCCTTGTCTTCATCCTT
 GTGGGTGGTAGTAGCAGGAAGTGTTCTGGCTGGATTTGCCTATAGTGTAGTCTTGACGAC
 GGTCTTTCAACTTGTCTCTGAACGAATTCAGCTAACTCCTCAATCAAGCAACTTCATT
 TGCTGTATTAGGCTGTAGTTTCGGAGCCTTTACGACCCCATTCGTTCTAGGTGCAATTGG
 CTTACTAACTCACAATGGGATGTTGGTCTTTAGTATCTTAGGAGGTTGGTTGATTGTAAT
 CTCTATCTTTGTCTATGTACCTACTTCAGAAGAGAGCTCTAGGATTGATTCCTAAGTTTTT
 CTTTTGATACTCAATGAAAATCAAAGAGCAAACCTATAGTTGATTGAGTTTGAATAGTAT
 GCTGTAG

ORF Predictions:

ORF #	Start	End	Direction	Length
1	356	1027	F	224 aa

> 3860582-1 ORF translation from 356-1027, direction F
 VLPLYLLFVPYGKSKKEVKKRAKEASRLTREMKGFLIFTLAIEAAVVVCTNTAITIRIPSL

MVERGLGDAQLSSFVLSIMQLIGIVAGVSFSFLISIFKEKLLLWSGITFGLGQIVIALSS
 SLWVVVAGSVLAGFAYSVVLTTVFQLVSERIPAKLLNQATSFVAVLGCSFGAFTTTPFVLGA
 IGLLTHNGMLVFSILGGWLIVISIFVMYLLQKRALGLIPKFFF*

Description:

unknown

Assembly ID: 3860724

Assembly Length: 1191bp

> 3860724 Strep Assembly -- Assembly id#3860724

GGATTCCAACGATTATGAACTTGACTGGTCCACTGATTCATCCAATGGCTTTAGAAACAC
 AGCTTTCTTGGAATTAGTCGTCCAGACTCCTAGAAAGTACAGCTCAGGTTTTGAAAATAT
 GGTGCGAAACGTGCCATCGTGGTTGCTGGACCAGAAGGGTTGGATGAAGCTGGCTTGAAC
 GGAACAACCNAGATTGCACCTTNTTGAAAATGGCGAAATCAGCTTGTCAAGCTTTACTCCA
 GAGGATTTGGGAATGGAAGGCTATGCTATGGAAGATATTCGTGGTGGGAATGCTCAGGAA
 AATGCAGAAATTTTGCTTAGCGTTCTGAAAAACGAAGCAAGTCCATTCTTGGAACGACA
 GTCTTGAATGCTGGTCTTGGTTTCTATGCTAATGGTAAGATTGATAGCATCAAGGAAGGA
 GTTGCCTTGGCCCGTCAAGTGATTGCTAGAGGCAAGGCCCTTGAAAACTCAGACTGTTA
 CAGGAGTACCAAAAATGAGTCAGGAATTTTGTAGCACGAATCTTAGAGCAGAAGGCGCGTG
 AGGTGGAGCAGATGAAGCTGGAGCAAATCCAGCCTCTGCGCCAGACCTATCGCTTGGCAG
 AATTTTTGAAGAATCATCAGGACCGCTTGCAGGTAATCGCTGAGTCAAGAAAGCTAGCCC
 TAGTTTGGGAGATATCAATCTCGATGTGGATATTGTGCAACAGGCCCAGACTTATGAAGA
 AAACGGAGCAGTGATGATTTCTGGTGTGACAGATGAGGTTTTCTTTAAAGGGCATTTGGA
 TTATCTACGGGAAATTTCCAGTCAGGTAGAGATTCCGACGCTCAACAAAGACTTTATCAT
 AGATGAAAAGCAAATCATCCGCGCTCGCAATGCAGGTGCGACAGTTATCTTGCTTATTGT
 GGCAGCCTTGTCCGAAGAACGCCTCAAGGAAGTGTATGACTACGCGACAGAGCTTGGTCT
 GGAAGTCTTAGTGGAGACTCACAATCTAGCTGAACTAGAGGTAGCCACAGACTTGGTGG
 CTGAGATTATCGGGGTCAACAACCGCAACTTGACTACCTTTGAAGTCGACTTGCAGACCA
 GTGTAGATTTAGCCCCTTACTTTGAGGAAGGTCGCTATTACATTTCTGAATCTGCCATTT
 TCACAGGGCAGGATGCGGAACGACTAGCCCCATACTTTAACGGAATTCGAT

ORF Predictions:

ORF #	Start	End	Direction	Length
1	139	498	F	120 aa
2	686	1024	F	113 aa

> 3860724-1 ORF translation from 139-498, direction F

VVAGPEGLDEAGLNGTTXIALXENGEISLSSFTPEDLGMGYAMEDIRGGNAQENAEILL
 SVLKNEASPFLETTVLNAGLGFYANGKIDSIKEGVALARQVIARGKALEKLRLQLQEYQK*

Description:

ANTHRANILATE PHOSPHORIBOSYLTRANSFERASE (EC 2.4.2.18). -
 LACTOCOCCUS LACTIS (SUB SP. LACTIS) (STREPTOCOCCUS LACTIS).

> 3860724-2 ORF translation from 686-1024, direction F
 VDIVQQAQTYEENGAVMISVLTDEVFFKGHLDYLREISSQVEIPTLNKDFIIDEKQIIRA
 RNAGATVILLIVAALSEERLKELYDYATELGLEVLVETHNLAELEVAHRLGG*

Description:

INDOLE-3-GLYCEROL PHOSPHATE SYNTHASE (EC 4.1.1.48) (IGPS). -
 LACTOCOCCUS LACTIS (SUBSP. LACTIS) (STREPTOCOCCUS LACTIS).

Assembly ID: 3860858

Assembly Length: 858bp

> 3860858 Strep Assembly -- Assembly id#3860858
 ATCGAATTTGCCAACCAAGAAAAATATCCCTTGGATGGTTCTTGGCAATGCAAGCAATAT
 CATCGTTTCGTGATGGTGGGATTCGTGGATTTGTCATCTTGTGTGACAAGCTCAATAACGT
 TTCTGTTGATGGCTATAACATTGAAGCAGAAGCTGGGGCTAACTTGATTGAAACAACCTCG
 CATTGCCCTCCGTCATAGTTTAACTGGCTTTGAGTTTGCTTGTGGTATTCCAGGAAGCGT
 TGGCGGTGCTGTCTTTATGAATGCGGGTGCCTATGGTGGCGAGATTGCTCACATCTTGCA
 GTCTTGTAAGGTCTTGACCAAGGATGGAGAAATCGAAACCCTGTCTGCTAAAGACTTGGC
 TTTTGGTTACCGCCATTCAGCTATTCAGGAGTCTGGTGCAGTTGTCTTGTGAGTTAAATT
 TGCCCTAGCTCCAGGAACCCATCAGGTTATCAAGCAGGAAATGGACCGCTTGACGCACCT
 ACGTGAACCTCAAGCAACCTTTGGAATACCCATCTTGTGGCTCGGTCTTTAAGCGTCCAGT
 CGGGCATTTTGCAGGTCAGTTCGAATTTTCAAGAGCTGGCTTGAAAGGCTATCGTATCGGT
 GGCGTAGAAGTGTGAGAAAAGCATGCAGGATTTATGATCAATGTGCGAGATGGAACGGCC
 AAAGACTACGAGGACTTGATCCAATCGGTTATCGAAAAAGTCAAGGAACACTCAGGTATT
 ACGCTTGAAAGAGAAGTCCGGATCTTGGGTGAAAAGCCTATCGGTAGCGAAGATGTATGCA
 GGTGGTTTTACTCCCTGCAAGAGGTAGTGGGGACCTGACAGAGCCCCGATCGGTAAATCT
 ATGAAAAAGAAGGAATTT

ORF Predictions:

ORF #	Start	End	Direction	Length
1	610	807	F	66 aa

> 3860858-1 ORF translation from 610-807, direction F
 VSEKHAGFMINVADGTAKDYEDLIQSVIEKVKEHSGITLEREVRLGESLSVAKMYAGGF
 TPCKR*

Description:

unknown

Assembly ID: 3860890
 Assembly Length: 980bp

> 3860890 Strep Assembly -- Assembly id#3860890

CTGAAAAAACAGGTTTGGACTATGNAGATTGACAGACGACCGTTCGGAGGTGCAGATATT
 GATGCAGCAGGACCTCCCTTACCTGATGAAACCCTTAAGGCAAGTAGGGAAGCAGATGCT
 ATCCTACTAGTAGCTATCGGTAGTCCTCAGTATGATGGAGTAGCGGTTTCGCCCTGAACAA
 GGCCTGATGGCTCTCCGTAAGAACTCAATCTTTACGCTAATATTCGTCCTGTAAAAATCT
 TTGACAGTCTCAAGTATTTGTCACCACTCAAACCGGAACGAATTTCTGGTGTAGACTTCG
 TCGTGGTGCCTGAATTGACTAGGCGAGATTTACTTTGGAGATCATATCCTTGAAGAGCGC
 AAAGCGCGTGATATCAACGACTATAGCTATGAGGAAGTGGAGCGGATTATTTCGCAAAGCC
 TTTGCCATCGAATTGCAAGAAATCGCAGAAAAATCGTTACTAGTATCGATAAGCAAAATG
 TTCTAGCGACCTCAAACTCTGGCGGAAAGTAGCTGAGGAAGTCGCACAGGATTTCTCAG
 ATGTAACCTTGAACACCAGCTGGTAGACTCAGCTGCTATGCTTATGATTACCAATCCTG
 CTAAGTTTGATGTTATTGTAACGGAGAATCTTTTTTGGAGATATTTTATCTGATGAATCAA
 GCGTCTTATCTGGTACACTTGGGGTTATGCCATCAGCCAGTCATTCTGAAAATGGACCAA
 GTCTCTATGAACCTATTTCACGGTTCAGCACCTGATATTGCAGGTCAAGGAATTGCCAATC
 CTATTTCCATGATTTTATCAGTTGTCATGATGTTGAGAGATAGTTTCGGACGTTATGAGG
 ATACAGAGCGTATCAAACGTGCTGTTGAGACAAGTCTGGCGGCAGGAATTTTAACGAGAG
 ATATAGGAGGTCAGGCTTCAACAAAGGAAATGATGGAAGCTATTATTGCAAGGTTATGAA
 GTTAGACGAAAAAATTCGAT

ORF Predictions:

ORF #	Start	End	Direction	Length
1	397	486	F	30 aa

> 3860890-2 ORF translation from 397-486, direction F
 VERIIRKAFAIELQEIAEKSLVLSISKMF*

Description:

unknown

Assembly ID: 3860952
 Assembly Length: 874bp

> 3860952 Strep Assembly -- Assembly id#3860952

TCGATCTAGAGAATTGCTCCAGAGCTTCCTGACCGTCCGCTGCCTCAATAGTTTCATAGC
 CACAATCCGTCAAATAATCACTGACCCCTCACGGATCATCTCTTCATCTTCTACAATTA
 AAATTTTCATACTTTAACTGCTCTCTATTTTTTATTTTTCTTAGAATAAATACCTACTCT
 ATTTTCTATTATAGTCTCTTGCTGGCCTTTTGTATGTAAGCAACTGACCACTAGATAAAA
 CGTTGTGAAATTCCTTTCTCATAAATTCCATAACTTTAGTATATTATATTTAAGCACTAA
 AGTACAAAGAAAGCAACTGAAAGCAATGATTTTACCACCTGCTTTCAGATTTATTTTGAA

TTGTTAAATAGCTATTCCTATCCACTATTCTTGAATAGAAACACAAGATGCAATCTTTAT
TCCAGACTCATTTTTTTAAAAAATCAAATTTATTACCATCCAGCAAGAGCTCTTTTGTT
GTTTTCTAAGGAGATTGCTTGAAGCAAGCGCCATAACGAGAACCACTAGAACCAAGGCAA
GGACAAAAATGATGATAAAGTCTGATGTCTGAATGGAAATGTCTAGGCTCGACAAGGTCT
TGCTAAAGCCATCTACTTCTGCACCGCCACCAAGGTTAGAGGCTTGAGCCGCCTTACTAG
CCTGTTTGGCAACACCTGAAGTCACATTGGCAAGGACAGTGTTTCCAATTCGCACGGGCA
GTGTAATTAGCTAGGAAGTAAGCANAACTAGAGCAGGGATAGCAATCAAGATAGATTCTG
GTGATGAATTGACCCAAGATACTTGCCTGCTTGAGACCAATAGAGAGGAGGATTCCCCT
TCCTTGCCGACGGGCATTGATCCAAAGACTGAGC

ORF Predictions:

ORF #	Start	End	Direction	Length
1	449	715	R	89 aa

> 3860952-1 ORF translation from 449-715, direction R
VRIGNTVLANVTSGVAKQASKAAQASNLGGGAEVDGFSKTLSSLDISIQTSDFIIFVLA
LVLVVLVMALASSNLLRKQPKELLLDGE*

Description:

unknown

Assembly ID: 3860962

Assembly Length: 762bp

> 3860962 Strep Assembly -- Assembly id#3860962

CTTGTAACGGTCATAAAGTTTCTGCAAACTACCATCCTTGCTCCATTTAGTAACCAAGTT
ATCAAGATAGTCGTTGAGCTCTGTATTTGATTTCTTGGTAACAATACCGTAGTCAGATGG
CTTGAAACTATCATCTAGTAGTTCTGTGCGTTTAACTAGTGTAGCCAGATAGAATAGAGC
GGTCAACGGAAAAGGCATCGATACGATGAGCGTGAAGGGAAGTAATCAATTCTGGGTAGG
AACCAAGTTTCGACGAATTTAAACTTCAGACCTTTCTTTTACCCAGTTTCAGTAATCAGGC
GTTGGGTGATAGAACCCTTGGGCGACTCCGATGGTTTTGCGTTTAGGTCCCTCAATCTTTT
TGATTTTGGCAGATTTATTGACCAAAAATCCAGAAGCGTCTGTGTAGTAGGGACTGGTAA
AGTTGTAGAGTTTTTTGCGTTCGTCCGTGATGGTAAAGGTCGCGATATCCATATCGACCT
GTTTCATTGTCTAGAAGGGGGCCGCGGGTTTGTGCTGTAACCGGCACATAGTGAATCTTGA
CCTTGAGTTTCATCAGCTACCATTTTGGCCAAGTCGGTTTCGATACCAGAATAAGTACCGG
TCTTGGGATCTTTGTTAACCAAAATTTGGGAACGTCTTGTTTGACACCCGACAACCAGTTC
GCCTCTTTTTTTGAATGTCTGCGATACTAGTATTAGCCTGGACTGGTTTGGCAGCAACAAG
GCCGAAAAGGCTAATCAATAATGCTGATAAAAAGAATTCGAT

ORF Predictions:

ORF #	Start	End	Direction	Length

1 152 646 R 165 aa

> 3860962-1 ORF translation from 152-646, direction R
VSNKTFPILVNKDPKTGTYSGIETDLAKMVADELKVKIHYVPVTAQTRGPLLDNEQVDM
IATFTITDERKKLYNFTSPYYTDASGFLVNKSAKIKKIEDLNGKTIGVAQGSITQRLITE
LGKKKGLKFKFVELGSYPELITSLHAHRIDAFSVDRSILSGYTS*

Description:

cell adhesion factor PEB1 precursor - Campylobacter jejuni

Assembly ID: 3861268

Assembly Length: 1942bp

> 3861268 Strep Assembly -- Assembly id#3861268
CTCGAATTTTTGGTGCTCCAGAAACGGTTCAGCAGGAAGCGTTGCTTTCAAGGCATCCA
TGGCAGTGAGTTCTGCAAGCAAACGTCCCTTGACCACACTGGTCAAATGCATGACGTAGC
GGAAGAGCTCCACCTCCATATACTTAGTAACTTGGACACTGGCCGTTTCAGAGATGCGGC
CAATATCGTTACGCCCCAAGTCTACCAACATTCGATGTTCTGCTGTTTCTTCTCATCAG
AGAGGAGGTCAGTCGCCAAGGCCTTGTCTTCTCCATCCGTAGCCCCCTCTTGGTCGCGTCC
CTGCAATCGGATTGGTTGTCACGATGCCATTTTTGACAGAAACCAAACCTTCTGGACTAG
CTCCGATGATTTGATAATCCCCAAAATCATACAAATAAAGGTAATTAGATGGATTAGTCA
CGCGGAGATTTCTGTAGAAGTCAAATGGATTTCCAGTTAACTTCTGCGTGAAGAAAACGC
TGGCTGAGTTACACATCGGAACATATCTCCGTTACGAATCAAGTCACGAGCTGTTTCTAC
CATTCCTCAAACCTTATGTGGAGCGATATGCGGTTTGAAGTCAAGTGGTGATAAATCCAA
GTCTTCAAATTCATTTGGAGCAGGAATGCGTAATTCCTCAAGCACTTGGTTCAAGGATTT
TTCCAAGGCCTCTTGACTGCGCTCACTATAAAGTGCATCCTCTATGACATGTTATCTTCT
CCTTCTTGTTGGTCAAAGACCATATAGCTCTCATAGACAAAGAAATGCATGTGCGGCGTC
CCAATTGTATCCTCAGGGATTTGACCAATTTCTTTCATAAAGCGAAATCATATCGTAACCA
ACAAAACCAATGGCTCCCCCACCAAAAGGGAGGTCTGAATGGTGCTGGCTCTTATGAATC
ACTTCATAAAGGAAATCCAAGGGATCCCGATCAATCGCTTGACCATTTTGATAGAGAACT
CCATTTTCAAACCTTAATCTCAAAAACCTGGATTATAGGCTAGGATAGAAAAACGAGCTGTT
TCCTTGCTCTCTCGGAATACTCTCTAAAATAACCTTATGTTGCCCCTTTAAGCGCATATAA
GCCAAGATTGGTGATAAGACATCTCCATGAATGATTTCGTTCCATTGTCATTTCCCTTTCA
GTTCTAATTCGAGTTCGTGGCGACTGTATGAAAAATCCCCACGCAAAATAACTTGCGTGA
GGACGAAATTCGCGGTGCCACCTCAATTATAGGATTTCTCCTATCTCTCATTCCTGTCTC
AGATATCTCCTGTAAACAGGCTGTGCGATAAAGGGCACTCCCTTGAGAATGATGTTTTCTT
CTCTCGTTTCAGATGAACCCAACCTTTACAGCTTTCTCTGCTTGTTTTTCAGCAACCACAAG
CTCTCTGTGAGAGAAAAGACTGTAATTTTTTCCATCTATTATTTTTTAGCTTCTAGTAATC
TGCAATCGCAGCTAGGTCCTTGCCCTCCACGACCAGAGACATTGATGAAGAGATGTTTCATC
TCGGTACACCTTTATACTCTTCGAAAATCTCTTCAAACCGCGTCAACGTCGCTTGCCGT
AGGTATGGTTACTGACTTCGTCACTTCTATCTGCAACCTCAAACAGTGTTTTGAGCTGA
CTTCGTCAGTCTTATCGACAACCTCAAACAGTGTTTTGAGCAGCCTGCAGCTAGTTTCC
TAGTTTGCTCTTTGATTTTCATTGAGTATTATTTTCATTTTCTCCTGCAATTGAATTCCTG

CTCAGCTTTTTGTCTTCTATTTCTTTAAAATCAAAGTAGCTCTTTTGTTAATAACTCGAT
 CAACAAACATCGTGGTACAAGTATCTACTTTGAAATTTATCAACCACTTAACAACGATA
 CTGTATTTCTAGGAAAACGATGACATTCTTCCTAATAAACTTCTCATATATAGCATAAA
 TTTCTACTCTTTTAAATTCGAT

ORF Predictions:

ORF #	Start	End	Direction	Length
1	457	645	R	63 aa

> 3861268-1 ORF translation from 457-645, direction R
 VLEELRIPAPNEFEDLDLSPLDFKPHIAPHKFEGMVETARDLIRNGDMFRCVTQPAFSSR
 RS*

Description:

ANTHRANILATE SYNTHASE COMPONENT I (EC 4.1.3.27). - LACTOCOCCUS
 LACTIS (SUBSP. L ACTIS) (STREPTOCOCCUS LACTIS).

Assembly ID: 3861270

Assembly Length: 1048bp

> 3861270 Strep Assembly -- Assembly id#3861270

CTGTTAAGATTGTTTCCGTGCATCCACATAGGATTTACCTTGTCTGTATGGGCCAATTCA
 CCCATCAAAACGCCATAGGTCTCATCTGTCAAGATACTAGACATACCGATATTGTACCAA
 AGACTGGTATGACGGAATAAGTCGATGCGTGTAACCTCAACAAAAAGAGACGCAAGTTG
 ATTAGAAAAACCGTCATAGCAATAGCTGCCACAGGAGCTTGAACCACAATCAGTGCCAAC
 ATGGCAAACCTGGGCACTCCCAGCATAAACAAAGAGACTCATCAAGCCATCTCAACAGGT
 GTCACATAGGGCGCACCGATAGTCCCACAGGCCAGGCCGATACTGACATAGCCAAGAGCC
 GTTGGCATGGCTGCCTGCGCCCCCTCCTAAAATCCTTTTTCTTTTCATCTTTCTCCTCATA
 TTGTCTTAATAATACTCAATGAAAATCAAAGAGCAAACCTAGGAAATTAGCCGCAGGNTGC
 TCAAAACACCGTTTTGAGGTTGCAGATAGAACTGACGAAGTCAGCTCAAAACACCGTTT
 TGAGGTTGCAGATAGAACTGACGAAGTCAGTAACATATATACGGCAAGGCGACGTTGACG
 TGGTTTGAAGAGATTTTCGAAGAGTATTAGAAAATGCCGATAAGGGTCTGCATACCAAGG
 CTGGTGAGGATGATGGCAATCCAGCAGACGGCTCCGAGAACAATGGATTTTCCACTGGAT
 TTGACCATAGCGACCAGATTAGTTTTGAGACCGATGGCACTCATGGCCATGATAATGAGG
 AATTTAGAGAGTTGTTTGAGAGGGGTAAAGAACTACTAGACACACCGAGAGAGGTCAGA
 AGGGTGGTTAGGAGCGATGCAAGGATGAAGTAAAGGATAAAAAGTGGAAGACTTTTTTTC
 AGTTGTAAGCCTTGCTTATTTTTTTTGCTCGCGACTTTGCCAGTAGGAGAGAAAGAGAGTG
 ATGGGGATGATAGCTAGGGTGCGCGTGAGTTTGACAATGGTTGCGGATTTCGAGGGTATTG
 GTCTGGTAGAGACTGTCCCAAGCGCTAG

ORF Predictions:

ORF #	Start	End	Direction	Length
-------	-------	-----	-----------	--------

1 627 824 R 66 aa

> 3861270-1 ORF translation from 627-824, direction R.

VSSSFFTPLKQLSKFLIIMAMSAIGLKTNLVAMVKSSGKSIVLGAVCWIAIILTSLGMQT
LIGIF*

Description:

unknown

Assembly ID: 3861288

Assembly Length: 1571bp

> 3861288 Strep Assembly -- Assembly id#3861288

AGAGCTGGTAATATTCCCAAAGAAACGGCTCAAATCGAATTAGAAAGCCTTCTGCAAAAA
GGAATCCCAGTCGCTCTGGTATCACGATGCTTTAACGGTATTGCCGAGCCTGTTTATGCC
TACCAGGGTGGGGGCGTACAGTTGCAAAAAGCAGGCGTTTTCTTTGTAAAGAACTCAAC
GCCCCAAAAGCCCGCTTGAAACTCCTCATCGCCCTCAATGCCGACTAACAGGACAGGCT
TTGAAAGACTATATGGAAGGCTAATACTCTTCGAAAATCTCTGCAAAACCACGTCAGCGTC
GCCTTACCGTATGTAGAGCACAAAATCAGGAAATCTTCTCGATTCCCTGATTTTTTCTAT
TTACGTTTTTCGTGTTGAGCTACGTTCTGTCAAACCATGAGGTAAGAGAACTTCACGTTCT
TCCAACCTCTTCCTTATGCATAATCTTGGTCAACATACGCATACTAATGGCACCAAGGTCA
TAAAGAGGTTGGGCAATCGTTGTCAAGTTTGGACGGGTAAAGCGTGAGATTTGTGAATCA
TCACTAGTAATAATTTCGATAATCTTCTGGCACAGAAACACCTTATCAGCCAAACCGTTCA
AGACTCCTGCTGCCAACTCATCACCTGTCACAACTGCTGCAGTTGCATTTGATGAAATCA
AACGCTCTGCTAAGGCGTAACCATCATCATAGCTATATTTAGATTCAAATACCAAACCTT
CACTATAAGCGATTCCCTGCTTTTTTCAAGGTTTCCTTG TAGCCAACTAAACGAACCTTAC
CATTGATGTCATCCACTAGCGGACCGCTAACGAAAGCAATACGCTCATTTTCTTTAGCAA
GGTAACCTCACTGCATCAATTGTTGCTTGCTTATAGTCAATATTGACACTTG GCAACTGGT
GCTCAACATCGACAGTTCCTGCGAGAACAATCGGAGTACGTGAACGCGAAAATTCTGAGC
GAATTTTATCTGTCAAGTGATAACCCATATAGATAATGCCATCTACCTGCTTTGAAAAGA
GGGTATTGACAACAGAACTTCTTTCTCGTTATCTTCATCGCTATTAGCTAGGACAATAT
TGTACTTGTAATTTCTGCAATATCATCAATCCCCCTTAGCCAAACTCGAAAAATAACCAT
TGGTAATATTTGGAATCACGACACCGACAGTGGTTGTCTTTTACTTGCAAGACCACGCG
CAACTGCATTTTGACGATAATCCAAACGATCAATTACCTCTAGCACTTTTTTACGGGTAT
TCTCTTTTACATTTTTTATTGCCATTGACCACACGGCTGACCGTCGCCATGGGAAACACCT
GCTTCACGAGCGACATCATAAATGGTTACTGTATCATCTGCATTCATTCCTTTTCTGTC
CTTTCTATCTCCACACATTCTTTTACAAGTAGAAGTGCTGAATTGAAAGCTCTATATCTT
ACTTACAAAAATGAAGATGTGAAAATTTTCGTTTTTCATATTTCTACTTATTCCATTCTATC
ACTAATTGTAAACACTTTCAAGTGTTTTTTGAAGATTGATTGAAAAAATTTCATAGAAAA
CCTAGGTTTAG

ORF Predictions:

ORF #	Start	End	Direction	Length
1	357	572	R	72 aa

> 3861288-1 ORF translation from 357-572, direction R
 VPEDYRIITSDDSQISRFRTRPNLTTIAQPLYDLGAISMRMLTKIMHKEELEEREVLLPHG
 LTERSSTRKRK*

Description:

GLUCOSE-RESISTANCE AMYLASE REGULATOR. - BACILLUS SUBTILIS.

Assembly ID: 3861306

Assembly Length: 1682bp

> 3861306 Strep Assembly -- Assembly id#3861306

CTGACGTAAAAAAGATTTTCGGAAAAGTATCATCATCTATTTTAGACCATTTTCTTATAA
 TAACCATTTTATTTTATTTGTCAAGGTCTTTGAATTCTTTCTTAAACAAGCCTTGTAAT
 CTCTACTTTTGAAGAATTTATTTTTCCTTACTGACAAGATTTGAGACGGTAGGAATCATT
 GAAAATAACCTAGCCAACATCAATCACAATCATTCTCCTTTCTCAATTACACTAAATTA
 TAGTGTATTGAATCTATAACAGTGCACCTTGGCTGCTAAAATATTTCTATAAATTAATTT
 GACTTTCCTGATAGAGTTGTTACATCTTATTTCAATTCATACTTTCCCTTATACTC
 AATGAAAATCAAAGCGCAAACCTAGGAAGCTAGCCACAGGCTGCTCAAAGCACTGCTTTGA
 GGTGTAGATAAGACTGACGAAGTCAGTTACATATATCTACGGCAAGGCGAAGCTGACGC
 GGTTTGAAGAGATTTTCGAAGAGTATAAAGTTTGTTTCTGTATCTTTCAGAAAAATAAGG
 TATACTGTATGTAAACGATTTCAAAGGAGTCCAGTTATGGCAAAAACATTTTATTTATCCA
 AATAAACAGAGCATTTTAGGAGAACAAGAGATTTTGAATGCCAAGTCGATCTTGGCTATG
 ATGTAGTCTATCTCCGTCAGCCTCTTAATCGTCTCGAGTATATTGAGTGTGCGATAGTGG
 GGCAATCACAATTTCTTTTAAAGGTCAGTTATGCTGATGGTCAAAGGCTTACCGTGTCG
 ATCTTCCTGACCTACTAACAAGACAGACTGGCAGATTATCAAGTCATTTTATAGATGTTT
 TGCTTGCTTATACAGGGACTGATATTGAAGGGCTAGATGGTTTTGATTTTGAAGCTTATT
 TCCAAGCAAGTATTCAAGCCTATCTAGCAGACCCTGTAGCTCGTTTTACGATTTGCCAAC
 GAATTTTAAATCCTATTTTCTTTAGTCGTGAGAACTTGAAAAGCTTTTATAGAGGCAGATG
 GCTTGGCTCAGTTTGAAGCGCGTGTGCGTGCGGTTCAAGAGACAGATGCCTACTTTGCGA
 GAGTTTCCTTCTATCAGGATGGAGAAGGAAAAGTGCATGGCGTTTACCATCTAGCTCAAG
 GAGTCAAGACAGTTTACCGAGAGAACCGTTTGTTCCTGCAGCCTATATTGAGCGAATTG
 GTGGATAAGGAAGTCCAGTGGGAGATTGACTTGGTTCAAATCACAGGAGACGGCTCTAAA
 CCAGAAGACTATGAATCCATAGCTCGCTTGGACTATGCAAAATTTCTTAGAGGTATTACCC
 CCATCTTTTACCACCAACTAGACGCCAATCAAATAGAAATACAACCCATCCTAGGACAA
 GATTTTAAACATTAGCACAAGAAAAGTAAAGCAGAAGCAGGTCAATCGACTTGCTTTTTT
 TGACATAGAAAAATCCTGCCAAGGATGACAGGATTGCTACTCAATGAAAATCAAAGAGC
 AAACCTAGGAAGCTAGCCGCAGGCTGTACTTGAGTACGGTAAGGCGAAGCTGACGTGGTTT
 GAATTTGATTTTTCGAAGAGTATGAATTTTAAAGAAAGGCCAAGATACGAAGATAATCTCC
 AATCAGTGCCACTTCAGCTTCCAAGAAGAAGAAGATTATAACTCCCGTTCCCCCAAGGACA

GA

ORF Predictions:

ORF #	Start	End	Direction	Length
1	717	1208	F	164 aa
2	1201	1410	F	70 aa

> 3861306-1 ORF translation from 717-1208, direction F
 VGQSQFLFKVSYADGQKAYRVDLPDLLTKTDWQIIKSFLDVLLAYTGTDIEGLDGFDFEA
 YFQASIQAYLADPVARFTICQRIFNPIFFSRENLSKFLEADGLAQFEARVRAVQETDAYF
 ARVSFYQDGEKGVHGVYHLAQGVKTVLPREPFVPAAYIERIGG*

Description:

unknown

> 3861306-2 ORF translation from 1201-1410, direction F
 VDKEVQWEIDLTVQITGDGSKPEDYESIARLDYAKFLEVLPPSFYHQLDANQIEIQPILGQ
 DFKTLAQEK*

Description:

unknown

Assembly ID: 3861334
 Assembly Length: 3041bp

> 3861334 Strep Assembly -- Assembly id#3861334
 ATCGAATTAAAAATGAGGTATTCAGGCTTGTGATTTTCTATGGAAGTTAATAGTGATTGC
 CTCTAATGCTTACAAGTGATATTAAAAATAGAGGACCTAGTGATGTCAATCATTTCAACT
 GATTTAACCCCTTTTCAAATAGATGATACATTGAAAGCAGCCTTGCGAGAAGATGTTTCAT
 TCCGAAGATTACAGTACCAATGCCATTTTGTATCATCATGGCCAAGCCAAGGTGTCGCTT
 TTTGCCAAGGAAGCTGGTGTTTTAGCGGGGCTAACCGTTTTTCAAAGGGTTTTTACCCTA
 TTTGATGCCGAGGTGACCTTCCAGAATCCTCATCAATTTAAGGATGGGGATCGTTTGACT
 AGTGGCGATTTGGTTTTAGAAATCATAGGCTCGGTGAGAAGTCTCTTAACATGTGAACGC
 GTTGCCCTTGAATTTTTTACAACATTTATCAGGGATCGCTTCGATGACAGCTGCTTATGTA
 GAAGCCTTAGGCGATGATTGCATTAAGGTATTTGATACTCGAAAACTACTCCTAATTTA
 CGTCTTTTTTGAGAAATATGCCGTGAGAGTTGGCGGTGGCTATAATCATCGCTTTAATTTA
 TCAGATGCTATCCTGCTAAAAGACAATCACATTGCGGCAGTAGGTAGTGTCAAAGGGCA
 ATTGCTCAAGCGCGTGCCTATGCTCCTTTTGTGAAAATGGTCGAGGTGGAAGTGGAAGC
 CTTGCTGCTGCCGAAGAAGCTGCGGCGGCGGGTGCTGATATTATCATGTTGGATAATATG
 TCATTGGAACAGATTGAACAGGCCATTACCCTAATTGCAGGACGTTCTCGGATTGAATGT
 TCTGGAAATATTGATATGACCACTATTAGCCGTTTTTCGTGGTTTTAGCGATTGATTACGTC
 TCCAGTGGTAGTTTAACCCATAGTGCTAAGAGTCTTGATTTTTCCATGAAGGGTTTAACC

TACCTTGATGTCTAAGTTGTAAAATAAACTAACTTTTTTAAAGGATGTCTTTCCTCTAGAA
 CGAGTTTTATGTCAGATAGTTTAAACGCCTCTTCAAATATAGTAAAATGAACCAAAAATA
 GTACACAATGTGGTATAATCTTCTTATGGCATATTTCAATAGATTTTCGTAAAAAAGTTCT
 TTCTTATTGTGAGCGAACAGGTAGTATAACAGAAGCATCACACGTTTTTCCAAATCTCACG
 TAATACCATTTATGGCTGGTTAAAGCTAAAAGAGAAAAACAGGAGAGCTAAACCACCAAGT
 AAAAGGAACAAAACCAAGAAAAGTTGATAGAGATAGACTTAAAACTATCTTACTGACAA
 TCCAGACGCTTATTTGACTGAAATAGCTTCTGAATTTGGCTGTCATCCAACCTACCATCCA
 CTATGCGCTCAAAGCTATGGGCTACACTCGAAAAAAGGACCACACCTACTATGAACAAGA
 CCCAGAAAAAGTAGCCTTATTTCTTAAAAATTTTAATAGTTTAAAGCACCTAGCACCTGT
 TTAGATTGATGAAACAGGATTTCGATACTTATTTTTTATCGAGAATATGGTCGCTCATTA
 AAGGTCAGTTAATAAGAGGTAAAGTATCTGGAAGAAGATATCAGAGGATTTCTTTGGTTGC
 AGGTCTAACAAATGGTGAGTTAATCGCTCCAATGACTTACGAAGAGACGATGACGAGCGA
 CTTTTTTGAAGCATGGTTTCAGAAGTTTCTCTTACCAACATTAACCACACCATCGGTTAT
 TATTATGGATAATGCAAGATTCCATAGAATGGGTAAGTTAGAAGTTTTATGCGAGGAGTT
 TGGGCATAAACTTTTACCTCTTCCCTCCCTACTCGCCTGAGTACAATCTTATTGAGAAAAC
 ATGGGCTCATATCAAAAAGCACCTCAAAAAGGTATTACCAAGTTGCAATACCTTTTATGA
 GGCTCTTTTGTCTGCTCTTGTTTCAATTGACTATAGTTTACGGATACAGTTGGGAAAGA
 AGTTAAATGTAGTTGGATTTCCACTAAAGGTTGATGAGTAAGTTTTTGTATCTGAACCTG
 ATTGGCCGCAAGCAGCTAAAAGCAAAGCAGATGCAAAAGTCAGACCTGCACCAAGGACAC
 GCTTCTTTATGTTTCATCTTCTTTCTCCTTAATAGTGGAATTTGTAAAGTTAATTGAATT
 TCAAGAATGAAGTTTTATAAACTTTGGTTATAAAAAACAAAGGATTTCTGTCTTTTATA
 CAGTCCTCCCCTTGTTTTTATACGATTTCAATTTTAAATTTTTCTGCAAAAAATATTTAT
 AGTAATTCCACACAGAAAGCATCCCATGGAAGTAAGATTTGTTTTTCAAAGACTTCTTGA
 GCTAGGGTGTTTTCAATCAAGACAGATTTGACTTTTCTTCTACTGTCAAGTCTTGCTCT
 TCATTGGACAAGTTAGCCACAAGTAGGAAGCGACGGTCGCCATCCTTACGTATATAAGCA
 AAGACCTTATCAGCCGTATCAAGCAATTCAAAGTCAGCTCGAATTAGCCAACCTATTCTCC
 TTGCGAATTTGGACCAGTTTCTGATAGGTATAGAAAATAGAATCTGGATTTGCCAGCGCT
 TCTTGGACGTTGATCATCTCGTAATTTGGATTAACTGCCAACCAAGGTTGACCTGTTGAG
 AAACCAGCGTTTTTGTCTCTCGTCCCATTGCATAGGGGTACGGGCATTGTCACGTCCAATA
 ACACGGATACTGTCCATGATTTCTTGTCATCGGAACACCTTTTTTCAAGAGCCTCACGCGCA
 TAGTTGAGAGATTCAATATCTTCTACTTGATCCAGTGTTTCAAACGGATAGTTGGTCATC
 CCAATCTCCTCACCTTGGTAGATATAAGGAGTTCCTCTCATAAGATGAAGCAAGATTGCA
 AAGGCTTTGGCAGATTTTTCGCGGTATTCTTGGTCATTTCCCCAGATTGAGACAATACGA
 GGGAGGTCATGGTTGTTCCAGAAGAGGGAATTCCAGCCGTCCTCAACTCCTAACTCTGTC
 TGCCATTTGTTGAAGATTTCTTTTAACTTAGCGATATTCAG

ORF Predictions:

ORF #	Start	End	Direction	Length
1	76	975	F	300 aa

> 3861334-1 ORF translation from 76-975, direction F

VILKIEDLVMSIISTDLTPFQIDDTLKAALREDVHSEDYSTNAIFDHHGQAKVSLFAKEA

GVLAGLTVFQRVFTLFD AEVTFQNP HQFKDGDRLTSGDLVLEIIIGSVRSLLTCERVALNF
LQHLSGSIASMTAAYVEALGDDCIKVFDTKRKTPNLRLFEKYAVRVGGGYNHRFNLSDAIL
LKDNHIAAVGVSQRAIAQARAYAPFVKMVEVEVESLAAAEAAAAGADIIMLDNMSLEQI
EQAITLIAGRSRIECSGNIDMTTISRFRGLAIDYVSSGSLTHSAKSLDFSMKGLTYLDV*

Description:

PROBABLE NICOTINATE-NUCLEOTIDE PYROPHOSPHORYLASE (CARBOXYLATING)
(EC 2.4.2.19) (QUINOLINATE PHOSPHORIBOSYLTRANSFERASE
(DECARBOXYLATING)) (QAPRTASE) (FRAGMENT). - BACILLUS
SUBTILIS (BLAST)

Assembly ID: 3864148

Assembly Length: 4694bp

> 3864148 Strep Assembly -- Assembly id#3864148

TTAATTTAAATTCCTTAAATTTTTTTCATAATAATCTCCCTATAAAAATAAAGTCGCCCCAA
TCAGGCGGCTTATTTTTTTTGAAAAATGGGCTTGGTGCCTGAGAATAAATAGCTTAGTGAT
AGAAGAAAATGGGGAAATATGGTATAATGAAACGATAGATTTTTGAATAGGAATAAGATC
ATGTTTGGATTTTTTAAGAAAGATAAAGGCTGTGGAAGTAGAGGTTCCGACACAGGTTCC
TGCTCATATCGGCATCATCATGGATGGCAATGGCCGTTGGGCTAAAAACGTATGCAACC
GCGAGTTTTTTGGACATAAGGCGGGCATGGAAGCATTGCAAACCGTGACCAAGGCAGCCAA
CAAACCTGGGCGTCAAGGTTATTACGGTCTATGCTTTTTCTACGGAAAACCTGGACCCGTCC
AGATCAGGAAGTCAAGTTTATCATGAACTTGCCAGTAGAGTTTTATGATAATTATGTCCC
GGAACCTACATGCGAATAATGTTAAGATTCAAATGATTGGGGAGACAGACCGCCTGCCTAA
GCAAACCTTCGAAGCTTTAACCAAGGCTGAGGAATTGACTAAGAACAACACAGGATTGAT
TCTTAATTTTGCTCTTAACCTATGGTGGACGTGCTGAGATTACACAGGCGCTTAAGTTGAT
TTCCCAGGATGTTTTAGATGCCAAAATCAACCCAGGTGACATCACAGAGGAATTGATTGG
TAACTATCTCTTTACCCAGCATTTCGCTAAGGACTTACGAGACCCAGACTTGATTATCCG
TACTAGTGGAGAATTGCGTTTGAGCAATTTCCCTTCATGGCAGGGAGCCTATAGTGAGCT
TTATTTTACGGACACCTTATGGCCTGATTTTGACGAAGCGGCCTTGACAGGAAGCTATTCT
TGCCTATAATCGTCGCCATCGCCGATTTGGAGGAGTTTAGGAGGAAATATGACCCAGGAT
TTACAGAAAAGAACCTTGTTATGCAGGGATTGCCCTGACTATTTTCCTACCAATTTTAAT
GATTGGGGGCTCTTGCTTCAGATAGCAATCGGAATCATANCCATGCTAGCCATGCATGAA
CTTTTGAAGATGAGAGGTCTAGAGACCATGACGATGGAGGCCTCTTGACCCTCTTTGCAC
NTTNGTATTGACCATTCCCCCTGGAATCGAATTACCTGACTTTTTTGCCAGTTGATGGGAA
TGTGGTTGCCTATAGTGTTTTGATTTCAATCATGTTAGGAACGACCGTTTTTAGCAAGTC
TTATACGATTGAGGATGCGGTTTTCCCTCTTGCTATGAGCTTCTACGTGGGCTTTGGATT
TAATGCTTTTACTAGATGCTCGTGTGTCAGGTTTGGACAAGGCTCTCTTAGCCTTGTTGAT
CGTCTGGGCGACAGACAGTGGTGCCTATCTTGTTGGGATGAACTATGGGAAACGAAAGTT
AGCACCAAGGGTATCGCCTAATAAAACCCTTGAGGGTGCCTTGGGTGGTATTTTAGGAGC
AATTTTAGTAACCATTATCTTTATGATAGTTGACAGTACAGTTGCTCTTCATATGGAAT
TTACAAGATGTCAGTCTTTGCTATTTTCTTTAGCATTGCTGGACAATTTGGTGATTTACT
AGAAAGTTCGATCAAACGTCATTTTGGTGTTAAGGATTCTGGGAAATTTATCCCTGGACA

TGGTGGTGT TTTGGATCGTTTCGATAGTATGTTGCTTGTATTTCCAATCATGCACCTTAT
TGGACTCTTTTAATCAAAAGACGGAGGAAACGCTATGCTCGGAATTTTAACCTTTATTCT
GGTTTTTGGGATTATTGTAGTGGTGCACGAGTTCGGGCACTTCTACTTTGCCAAGAAATC
AGGGATTTTAGTACGTGAATTTGCCATCGGTATGGGACCTAAAATCTTTGCTCACATTGG
CAAGGATGGAACGGCCTATAACCATTCGAATCTTGCTCTGGGTGGCTATGTCCGCATGGC
CGGTTGGGGTGATGATACAACCTGAAATCAAGACAGGAACGCCTGTTAGTTTGACACTTGC
TGATGATGGTAAGGTAAACGCATCAATCTCTCAGGTAAAAAATTGGATCAAACAGCCCT
CCCTATGCAGGTGACCCAGTTTGATTTTGAAGACAAGCTCTTTATCAAAGGATTGGTTCT
GGAAGAAGAAAAACATTTGCAGTGGATCACGATGCAACGGTTGTGGAAGCAGATGGTAC
TGAGGTTTCGGATTGCACCTTTAGATGTTCAATATCAAAATGCGACTTTATCTGGGGCAAA
CTGATTACCAATTTTGCAGGTCCTATGAACAATTTTATCTTAGGTGTTGTTGTTTTTTGG
GTTTTAATCTTTATGCAGGGTGGTGTGAGAGATGTTGATACCAATCAGTTCATATCATG
CCCCAAGGTGCCTTGGCCAAGGTAGGAGTACCAGAAACGGCACAATTACCAAGATTGGC
TCACATGAGGTTAGCAACTGGGAAAGCTTGATCCAAGCTGTGGAACAGAAACCAAGAT
AAGACGGCACCGACTTTGGATGTGACTATTTCTGAAAAGGGGAGTGACAAACAAGTCACT
GTTACACCCGAAGATAGTCAAGGTCGTTACCTTCTAGGTGTTCAACCGGGGGTTAAGTCA
GATTTTCTATCCATGTTTGTAGGTGGTTTTACAACCTGCTGCTGACTCAGCTCTCCGAATT
CTCTCAGCTCTGAAAAATCTGATTTTCCAACCGGATTGAAACAAGTTGGGTGGACCTGTT
GCTATCTTTAAGGCAAGTAGTGATGCTGCTAAAAATGGAATTGAGAATATTCTTGTACTT
CTTGGAATGATTTCCATCAATATTGGGATTTTAAATCTTATTCCGATTCCAGCCTTGGA
TGGTGGTAAGATTGTGCTCAATATCCTAGAAGCCATCCGCCGCAAACCATTGAAACAAGA
AATTGAAACCTATGTCACCTTGGCCGGAGTGGTCATCATGGTTGTCTTGATGATTGCTGT
GACTTGGAATGACATTATGCGACTCTTTTTTAGATAATCGAGGAATATTATGAAACAAAG
TAAAATGCCTATCCCAACGCTTCGCGAAATGCCAAGCGATGCTCAAGTTATCAGCCATGC
TCTTATGTTGCGTGCTGGTTATGTTGCGCAAGTTTCAGCAGGTGTTTATTCTTATCTACC
ACTTGCCAACCGTGTGATTGAAAAAGCTAAAAACATCATGCGCCAAGAATTCGAAAAGAT
TGGTGCTGTTGAGATGTTGGCTCCAGCCCTTCTTAGTGCAGAATTGTGGCGTGAATCAGG
TCGTTACGAAACCTATGGTGAAGACCTTTACAACTGAAAAACCGTGAAAAATCAGACTT
TATCTTAGGTCCAACCTCACGAAGAAACCTTTACAGCTATTGTCCGTGATTCTGTAAATC
TTACAAGCAATTGCCACTCAACCTTTATCAAATTCAGCCCAAGTATCGTGATGAAAAACG
CCCACGTAATGGACTTCTTCGTACACGTGAGTTTATCATGAAGGATGCTTATAGTTTCCA
CGCTAACTATGATAGTTTGGATAGTGTATGATGAGTACAAAGCAGCCTATGAGCGTAT
TTTCACTCGTAGTGTTTAGACTTCAAGGCTATATTATTGGTGACGGTGGAGCCATGGGTGG
TAAGGATAGCCAAGAATTTATGGCCATTACATCTGCTCGTACAGACCTTGACCGCTGGGT
TGTCTTGACAAGTCAGTTGCCTCATTTGACGAAATTCCTGCAGAAGTGCAAGAAGAAAT
CAAGGCAGAATTGCTCAAATGGATAGTCTCTGGTGAAGATACCATTGCTTACTCAAGTGA
GTCTAGCTATGCAGCTAACTTAGAAATGGCAACAAACGAGTACAAACCAAGCAACCGTGT
TGTCGCTGAAGAAGAAGTTACTCGTGTTGAAACGCCAGATGTTAAATCAATTGATGAAGT
TGCAGCCTTCCTCAATGTTCCAGAAGAACAACGATTAAAACCTCTTCTACATTGCAGA
TGGTGAGCTTGTTGCAGCCCTTCTAGTTGGAAATGACCAACTCAACGAAGTCAAGTTGAA
AAATCACTTGGGAGCAAATTTCTTTGACGTTGCTAGCGAAGAAGAAGTGGCGAATGTTGT
TCAAGCAGGATTTGGTTCACCTTGGACCAGTTGGTTTGCCAGAGAATATTAAAATTTATTGC
AGATCGTAAGGTGCAAGATGTTGCGCAATGCAGTTGTGCGGTGCTAACGAAGATGGCTACCA

CTTGACTGGTGTGAACCCAGGCCGTGATTTTACTGCAGAATATGTGGATATCCGTGAAGT
 TCGTGAGGGTGAAATTTCCCCAGATGGACAAGGTGTCCTTAACCTTTGCGCGTGGTATTGA
 GATCGGTCATATTTTCAAACCTCGGAACTCGCTATTTCAGCAAGCATGGGAGCAGATGTCTT
 GGATGAAAATGGTCGTGCTGTGCCAATCATCATGGGATGTTACGGTATCGGTGTCAGCCG
 TCTTCTTTTCAGCAGTGATGGAGCAACACGCTCGCCTCTTTGTTAACAAAACGCCAAAAGG
 TGAATACCGTTACGCTTGGGGAATCAATTTCCCTAAAGAATTGGCACCATTGTGATGTGCA
 TTTGATTACTGTTAATGTCAAGGATGAAGAAGCGCAAGCCTTGACAGAAAACTTGAAGC
 AAGCTTGATGGGAG

ORF Predictions:

ORF #	Start	End	Direction	Length
1	212	940	F	243 aa
2	1202	1753	F	184 aa
3	2750	3037	F	96 aa

> 3864148-1 ORF translation from 212-940, direction F
 VEVEVPTQVPAHIGIIMDGNGRWAKKRMQPRVFGHKAGMEALQTVTKAANKLGVKVITVY
 AFSTENWTRPDQEVKFIMNLPVEFYDNYVPELHANNVKIQMIGETDRLPKQTFEALTKAE
 ELTKNNTGLILNFALNYGGRAEITQALKLISQDVLDKINPGDITEELIGNYLFQTQHLPK
 DLRDPDLIIRTSSELRLSNFLPWQGAYSELYFTDTLWPDFDEAALQEAILAYNRRHRRFG
 GV*

Description:

unknown

> 3864148-2 ORF translation from 1202-1753, direction F
 VVAYSVLISIMLGTTVFSKSYTIEDAVFPLAMSFYVGFGFNALLDARVAGLDKALLALCI
 VWATDSGAYLVGMNYGKRKLAPRVSPNKTLEGALGGILGAILVTIIFMIVDSTVALPYGI
 YKMSVFAIFFSIAGQFGDLLESSIKRHFGVKDSGKFIPGHGGVLDREFDSMLLVFPIMHLF
 GLF*

Description:

CDP-diglyceride synthetase (cdsA) homolog - Haemophilus
 influenzae (strain Rd K W20)

> 3864148-10 ORF translation from 2750-3037, direction F
 VDLLLLSLRQVVMMLLKMEFLRIFLYFLAMISINIGIFNLIPALDGGKIVLNILEAIRKRP
 LKQEIETYVTLAGVVIMVVLMIAVTWNDIMRLFRR*

Description:

unknown

Assembly ID: 3864172
 Assembly Length: 1352bp

```
> 3864172 Strep Assembly -- Assembly id#3864172
CTCGTAAGTTCGGAAGCTATCTACACAAGAAATTAACCGCTGCCTAAAGGAGAAGCCATG
TCAACATATAACTGGGATGAGAAGCATATCCTTACCTTTCCTGAAGAAAAAGTAGCCCTT
TCTACTAAGGATGTCCATGTTTACTATGGTAAAAATGAATCCATTAAGGGGATTGATATG
CAATTTGAAAGAAATAAAATTACAGCTTTGATTGGTCCGTCGGGATCGGGGAAATCTACC
TACTTACGCAGTCTCAATCGCATGAATGATACCATTGATATTGCTAAAGTAACTGGGCAG
ATTCTCTATCGTGGAATTGATGTCAACCGTCCAGAAATCAACGTTTATGAAATGCGTAAA
CACATTGGAATGGTTTTTCAACGCCCAATCCATTTGCTAAATCGAATTTACCGTAATAT
TACCTTTGCGCATGAACGTGCTGGAGTTAAGGATAAGCAAGTCCTAGATGAAATCGTAGA
AACCTCCCTTAGTCAGGCTGCCCTTTGGGATCAGGTTAAAGACGATCTCCACAAGTCAGC
CTTGACCTTATCAGGTGGTCAGCAACAACGTCTCTGTATCGCTCGTGCCATCTCTGTAA
GCCAGATATCCTCTTAATGGATGAGCCAGCCTCAGCCTTGGATCCGATTGCGACCATGCA
ACTAGAAGAGACCATGTTTGAGCTCAAGAAAACTTTACCATCATCATTGTAACGCATAA
TATGCAGCAGGCTGCTCGTGCAAGTGACTATACAGGCTTCTTTTACTTGGGTGATTTGAT
TGAGTATGACAAGACTGCAACTATTTTCCAAAATGCCAAGCTACAGTCCACCAATGACTA
TGTATCTGGTCACTTTTGGTTAGAAAGGAAACCGTATGACAGATGCGATTTTACAGGTATC
AGACCTGTCCGTTTATTATAATAAAAAGAAGGCTTTGAATAGTGTTTCCCTATCTTTCCA
ACCTAAGGAAATTACAGCCTTGATTGGTCCATCTGGATCAGGGAAGTCAACCCTCCTCAA
GTCTCTCAACCGCATGGGAGATCTCAATCCAGAGGTGACCACAACCTGGATCCGTGGTGTA
CAATGGTCAACAACATCTACAGTCCGCGTACAGATACGGTTGAATTACGTAAGGAAATCGG
AATGGTTTTTCCAACAACCTAATCCTTTCCCTATGACTATCTATGAGAATGTTGTCTACGG
GCTTCGTATCAATGGAATTAAGGATAAGCAGGTTCTGGATGAAGCCGTAGAAAAAGCCTT
GCAAGGTGCCTCTATCTGGGATGAGGTCAAGGATCGTCTATATGATTCAGCTATTGGATT
GTCAGGTGGTCAACAGCAGCGTGTCTGCGTGG
```

ORF Predictions:

ORF #	Start	End	Direction	Length
1	311	862	F	184 aa

```
> 3864172-2 ORF translation from 311-862, direction F
VELMSTVQKSTFMKCVNTLEWFFNAPIHLLNRIYRNITFAHERAGVKDKQVLDEIVETSL
SQAALWDQVKDDLHKSALTLSSGQQQRLCIARAI SVKPDILLMDEPASALDPIATMQLEE
TMFELKKNFTIIIVTHNMQQAARASDYTGFFYLGDLIEYDKTATIFQNAKLQSTNDYVSG
HFG*
```

Description:

HYPOTHETICAL ABC TRANSPORTER (ORF75). - BACILLUS SUBTILIS.
 (BLAST)

Assembly ID: 3864180
Assembly Length: 2258bp

> 3864180 Strep Assembly -- Assembly id#3864180
AACTTCGACCGTGATAAACAAGCTGAGCTTTGACATACTTGTAGCCAACCTAAAAGCCGT
TCTTCAAGGCCTCAAACCAGCTGCAACTCATTTCAGGAAGCCTGGATGAAAATGAAGTGGC
TGCCAATGTTGAAACCAGACCAGAACTCATCACAAGAACTGAAGAAATTCATTTGAAGT
TATCAAGAAAGAAAATCCTAATCCCAGCTGGTCAGGAAATATTATCACAGCAGGAGTCAA
AGGTGAACGAACTCATTACATCTCTGTACTCACTGAAAATGGAAAAACAACAGAAACAGT
CCTTGATAGCCAGGTAACCAAAGAAGTTATAAACCAAGTGGTTGAAGTTGGCGCTCCTGT
AACTCACAAGGGTGATGAAAGTGGTCTTGCACCAACTACTGAGGTAACCTAGACTGGA
TATCCAAGAAGAAGAAATTCATTTACCACAGTGACTCGTGAAAATCCACTCTTACTCAA
AGGAAAAACACAAGTCATTACTAAGGGTGTCAATGGACATCGTAGCAACTTCTACTCTGT
GAGCACTTCTGCCGATGGTAAGGAAGTGAACCACTTGTAATAAGTGTTCGTAGCACAGGA
AGCCGTTACTCAAATAGTCGAAGTCGGAAGTATGGTAACACATGTAGGCGATGAAAACGG
ACAAGCCGCTATTGCTGAAGAAAAACCAAACTAGAAATCCTAAGCCAACAGCTCCTGC
TGAGGAAAGCAAAGCTCTTCTCAAGATCCAGCTCCTGTGGTAATAGAGAAAAAACTTCC
TGAAACAGGAACTCACGATTCTGCAGGGACTAGTAGTCGCAGGACTCATGGCCACACTAG
CAGCCTATGGACTCACTAAAAGAAAAGAAGACTAAGTCTTTTCGATAAAAAATAAACAGC
GAGATTGAAGCTCGCTGTTTATTTTTTAATTAATCACCTAGTCCAAGACGTTCAAAGATA
TCATCCACTCGTTTGGTGTAATAAACTGGGTGAAGATTTTCATCGATTTCTTCTTGTGTG
AGACGTGATGTTACTTCTGAATCTGCCTCAAGAAGTGGTTTAAAGTCTACTTGGTTGTCC
CAAGAGTAGGCTGTTTTTGGTTGCACCAAGTCATAGGCTTGCTCACGGGTCATGCCTTTT
TCAATCAATGTCAACATAGCCCGTTGGCTAAAGATAAGACCAAAAGTCGAGTTCATGTTT
CGGATCATATTTTCTGGGAAGACTGTCAAGTTCCTTGACGATATTTCCAAAACGGTTGAGC
ATGTAGTCAATCAAATGGTTCGTATCTGGTGTGATGATACGCTCAGCTGATGAGTGAGAA
ATATCGCGTTCGTGCCAGAGAGCGACGTTTTTCATAAGCCGTAATCATGTGACCACGAATG
ACACGCGCCAGACCAGTCATATTTTCAGAACCGATTGGGTGCGTTTGTGAGGCATTGCT
GAAGACCCTTTTTGCCCCTTTAGCAAAGAACTCTTCTACTTCGCGTTGCTCAGATTTTTGT
AGACCACGAATCTCAGTCGCCATACGTTTCGATTGAAGTCGCAATGCTGGCAAGAACCGCA
AAGTACTCAGCGTGAAGGTCACGAGGAAGGACTTGTGTTAAAGATTCTTTGGGCACGGAT
GCCAAGATTTATCGCAGACATACTCCTCTACAAATGGTGGGATATTGGCAAAGTTCCCAA
CCGCACCAGAAATCTTACCAGCTTCTACACCAGCAGCCGCATGCTCGAAGCGCTCGATAT
TGCGTTTTCATTTTCGCTGTACCAAGTTGCTAATTTAAGACCAAAGGTTGTCGGCTCAGCGT
GCACACCATGAGTACGCCCCATCATGATGGTGAAGTTGTGCTCCTTGGCCTTGTGAGCGA
TGATATTAGTGAAGTTTTCAAGGTCACGACGGATGATGTCGTTGGCCTGCTTGTAGAGGT
AACCATAAGCAGTATCCACCACGTCGGTAGAAGTTAACCCTAGTGAACCCACTTGCCT
CTTCACCAAGAGTCTCAGAAACCGCACGCGTGAAAGCCACCACATCGTGGCGCGTCTCCT
GCTCAATTTCCAAAATACGGTCGATGTCAAAGTCCGCCTTCTTGCGAATCAAAGCCACAT
CTTCCTTAGGGATTTCCCCAACTCAGCCCATGCCTCGTCAGAGAGGATTTCCACCTCAA
GCCAAGCACGGTATTTATTTTCTTCACTCCAAATATTCGCCATCTCAGGGCGAGAGTAAC
GGTTGATCATGTGTTAATTTTTCTTCTTCTTAAGAT

ORF Predictions:

ORF #	Start	End	Direction	Length
1	930	1616	R	229 aa

> 3864180-2 ORF translation from 930-1616, direction R
 VPKESLTQVLPRDLHAIFYFAVLASIATSIERMATEIRGLQKSEQREVEEFFAKGQKGSSA
 MPHKRNPIGSENMTGLARVIRGHMITAYENVALWHERDISHSSAERIITPDTTILIDYML
 NRFGNIVKNLTVFPENMIRNMNSTFGLIFSQRAMLTLEKGMTREQAYDLVQPKTAYSWD
 NQVDFKPLLEADSEVTSRLTQEEIDEIFNPVYYTKRVDDIFERLGLGD*

Description:

ADENYLOSUCCINATE LYASE (EC 4.3.2.2) (ADENYLOSUCCINASE) (ASL). -
 BACILLUS SUBTIL IS.

Assembly ID: 3864184

Assembly Length: 4392bp

> 3864184 Strep Assembly -- Assembly id#3864184

CCCTTTTGCCCTCTCCCTTTGGTGCAGATTCTTTTGGGAATTGTGATTGGTCTCTTTTAC
 CCAATACTGACTTTCATCTTAATACGGAGTTGTTTTGGCCTGGTTATCGGACCCTTGCT
 TTTCCGAGAGGCTGAAGAAGCAGATGTTACGGCTATTTTAAACACTGGCGAATCATTTGT
 TTATCTCATATTTCCAGTGATTTTTATCTCGACCCTGAGTTTGGGTGGCTTGGCCCATCT
 TCTTTGGTTCAGCCTTCCCTTGGCAGCTTGCTTGGCTGTTGGGGCAGCCCTTGGTCCCTAC
 GGACTTGGTGGCCTTTGCCTCTCTTTTCGGAGCGTTTTAGCTTTCCTAAGCGCGTGTCCAA
 TATTCTTAAGGGCGAAGGACTCTTGAATGATGCTTCTGGTTTGGTGGCTTTTTCAGGTAGC
 TTTGACAGCTTGGACAACTGGAGCTTTTTCTCTGGGGCAAGCTAGCAGTTCGCTCATCTT
 TTCAATCCTAGGCGGTTTTTTAATTGGATTTTTAACAGCCATGACCAACCGCTTCCTCCA
 TACCTTCTTGCTAAGTGTGCGCGCAACGGATATTGCCAGTGAACTTTTATTAGAATTCTGA
 GTTTGCCTCTAGTGACCTTCTTCTGGCAGAAGAAGTCCATGTTTCAGGGATTATTGCCG
 TCGTAGTTGATCGAATTTTAAAGGCAAGTCGCTTCAAGAAAATCACGCTCCTCGAAGCCC
 AAGTGGATACGGTGACCGAGACGGTCTGGCATAACAGTGACCTTTATGCTCAACGGTCTG
 TCTTTGTGATTTTAGGGATGGAGTTGGAAATGATAGCAGAACCTATCTTGACCAATCCAA
 TCTATAATCCTCTACTTTTATTGCTATCTCTCATCGCCCTTACCTTTGTCCTCTTTGTCA
 TTCGTTTTTATTATGATCTATGGCTATTATGCCTATAGAACCCGACGCCTAAAGAAAAAGC
 TAAATAAGTATATGAAGGACATGTTTCTCTTGACCTTTTTCAGGTGTTAAGGGAACGGTGT
 CGATTGCTACGATTCTCTTGATACCAAGTAATCTAGAACAGGAGTATCCTCTCTTGCTTT
 TCCTTGTTGCAGGTGTGACGCTTGTGAGCTTTTTAACAGGTCTCTTGGTCTTGCCTCATC
 TTTCTGATGAAGAGGAAGAAAGCAAGGATTATCTCATGCATATCGCCATTTTGAATGAAG
 TAACGCTAGAGTTGGAAGAAAGAGTTGGAAGACACCAGAAATAAACTTCCCCTCTATGCGG
 CTATTGACAATTCGATCATGGACGTATTGAAAATCTCATTTTAAGCCAAGAAAACCAGGA
 TGATCAAGAAGACTGGGCTGCTTTGAAAATCGAATTCTTAGTATTGAAAGTGATGGTTTG
 GAACAGGCCTATGAAGAGGGGAACATTAGCAATCGTGCTTACCGAGTTTACCAACGTTAT

CTGAAAAATATAGAACAAGGAATCAATCGTAAACTTGCCTCAAGACTGACCTATTATTTT
CTTGTTTCCTTGAGGATTTTACGTTTTCTTCTTCATGAAGTTTTTACTCTTGGAAGACC
TTCCGTAGCTGGAAGGACAAGGAGCAAAGCCGTCTCCGTGCTCTTGATTATGACCAAATT
GCAGAGCTCTATCTTGCCAATACAGAGATGATTATTGAAAGTTTGGAAAACCTGAAGGGA
GTCTACAGACGCTCTTTGATTAGTTTTATGCAGGAGTCTCGTCTTCGAGAAACAGCTATT
ATCAGCAGTGGTGCCTTTGTGCAACGGGTTATCAATCGTGTCAAACCCAACAATATCGAT
GAAATGCTGAGAGGCTATTATCTGGAGCGCAAGTTGATTTTTCGAATACGAAGAAAAACGA
TTGATTACGACTAAGTATGCCAAGAAATTACGACAAAATGTAAATAACTTAGAGAACTAT
TCCTTGAAGGAAGCTGCCAATACCCTGCCGTATGATATGGTGGAATTGGTAAGAAGAAAT
TAGTTAATACTCTTCGAAAATCTCTTCAAACCACGTCAGCGTCGCCTTGGATTATATATG
TGACTGACTTCGTCACTTTCATCTACAACCTCAAAGCAGGGCTTTGAGCAACCTGCGGCT
AGCTTCCTAGTTTGCTCTTTGATTTTCATTGAGTATAAGATTGTAAGTGAAGGAGTGTGA
CATGAAAAAATGGGGAAAGAGCCTGAACTAGTCCTGTCTACTTTTACCCAATCACACTTC
CATTTGGTACAGCTGGATCAACTGTGAGAAGGGATCGAATTTGCCATCATGTTTCAGCTGA
GAGAATCATACCTGGCTGACATATTTTTTTCATCATTTTACGTGGTTTGAGGTTAGCAAC
GATTTGAACTTTCTTGCCGACCAATTCTTGTTCAATTTGGATAGTATTTTGCAATTCCTGA
AAGAATCTGACGATCTTCTCCATCACAGCATCCAAGCGGAATTGAAGCAACTTATCTGA
ACCTTCTACTTTAGACACTTCTTTGACTTCTGCGACACGGATTTCAACCTGTCAAAGTC
TTCAAACCTTGATTTTCATCCTTGTTTAGTTTGAGCTCAACTTCGTCCGGATTCCATTCCTT
TTCGACTGCTGGTTTTATTGCCTTCCATTTGTTCCCTTGATATAGGCGATTTCTTCTTCCAT
ATTTAGACGTGGAAAGATAGGTGTTTCTTTGGCAACTACAGTCACATCTGCTGGGAAGTC
AGCCAACTCAAGTTTTCAAGACTAGAACTTCTTCCAAACCAAGTTGAGTCAAACTGC
ACGACTAGTTTCCATCATAAATGGTTCAATCAAGTGAGCAACTACACGAATGCTGGCTGC
CAAGTGCTCATGACACTTGCCAATTGGTCAAGAGAGCTTCATCCTTGTCCAAGACCCA
TGGTGCAGTCTCATCGATGTATTTATTGGTACGAGAGATCAGAGTCCAGACTGCTTCAAG
CGCACGTGGATAGTCAACTGCTTCCATGTGTGTATGGAAGTCTGCGATTGATTTTTCTGC
AACCTCAGCAAGAACATGATCAAATTCAGTCACACCTTCTACATAGGCAGGGATTTGTCC
ATCAAAGTACTTATTAATCATGGAAACCGTACGGTTAAGGAGGTTCCCAAGGTCATTAGC
CAATTCATAGTTGATACGACCGACATAGTCTTCAGGAGTAAAGGTTCCGTCTGAACCAAC
TGGAAGGTTACGCATGAGGTAGTAACGAAGTGGATCTAGTCCATAACGCTCTACCAACAT
TTCAGGGTAAACGACATTCCCTTTTGACTTAGACATTTTTTCCGTCTTTCATGACAAACCA
ACCATGGGCAATCAAACGATCAGGTAATTTAACATCCAACATCATAAGAAGGATTGGCCA
GTAGATAGAGTGGAAGCGAAGGATGTCTTTTCCCTACCATATGGAAGACTGTTCCATTCCA
GAACTTGTCAAAGTTACCATGTTTCGTCTTGAGCGTAGCCAAAAGCTGTGCGATAGTTAAG
AAGGGCATCAATCCAAACGTAGACAACGTGTTTTGGATTTGATGGGACAGGCACTCCCCA
TGTAAGGTTGTACGAGATACCGCCAAATCTTCAAACCTGGCTCGATGAAGTTGCGTAG
CATTTTCATTAAAGACGACCATCTGGCGTGATAAATTCAGGATGAGCTTTGAAAAATTCGAC
CAAACGGTCTTGTTATTTGCTAAGGCGAAGGAAGTATGATTCTTCAGAAACCCATTCAAC
CTCATGACCTGATGGAGCAATACCACCAGTCACATTTCCAGCTTCATCACGGAAAACCTTC
TGCCAGCTGGCTTTCTGTAAAGAATTCTTCGTCTGATACTGAATACCAACCAGAGTATTC
ACCCAAGTAGATATCATCTTGAGCAAGTAAGCGTTCAAAGACCTGTGCGACAACCTTTTTC
ATGGTAGTCATCGGTTGTACGGATAAATTTATCGTATGAGATATCTAGTAATTGCCAGAG
TTCTTTAACTCCAACCGCCATTCCATCAACATAGGCTTGAGGTGTAATACCAGATTCGAA

TTCCGCTTTCTGCTGGATTTTCTGACCATGTTTCATCAAGACCTGTCAGATAAAATACATC
 GTAGCCCATCAGGCGTTTGTAACTGCTAGGACATCACATGCGATAGTTGTGTAGGCAGA
 ACCGATATGAAGTTTCCCAGATGGATAGTAAATCGGCGTTGTAATATAAAAATTTTTTTC
 AGACATAATTTTTCTTTCCAGGCAAATGAAACCTGTTTTTCTAACACTTCATTATATCA
 CATTTTTTAATGAATTCGATAGGGAAATCCATACCAAACAAGATAGACGAGTGTCCATC
 TTGTTGATCTCATTCATAACGAAGGGCTTCAATTGGATCAAGTTTCGATGCCTTGTGGC
 TGGCAAGACTCC

ORF Predictions:

ORF #	Start	End	Direction	Length
1	197	670	F	158 aa
2	612	1304	F	231 aa

> 3864184-1 ORF translation from 197-670, direction F
 VIFISTLSLGGLAHLLWFSPLAACLA VGAALGPTDLVAFASLSERFSF PKRVS NILKGE
 GLLNDASGLVAFQVALTAWTTGAFSLGQASSSLIFSILGGFLIGFLTAMTNRFLHTFLLS
 VRATDIASELLLEFEFASDDL SGRRS PCFRDYCRS*

Description:

unknown

> 3864184-2 ORF translation from 612-1304, direction F
 VTFFLAEEVHVSGIIAVVVDRILKASRFKKITLLEAQVDTVETVWHTVTFMLNGSVFVI
 LGMELEMI AEPI LTNPIYNPLLLLLSLIALTFVLVIRFIMIYGYAYRTRRLKKKL NKY
 MKDMFL LTFSGVKGT VSIATILLIPSNLEQEYPLLLFLVAGVTLV SFLTGLLVLP HLSDE
 EEESKDYL MHIAILNEVTLELEKELEDTRNKLPLYAAIDNSIMDV LKISF*

Description:

unknown

Assembly ID: 3864194

Assembly Length: 1941bp

> 3864194 Strep Assembly -- Assembly id#3864194

AATTAGTATTCTCAACCTTTTTATCTTGATAGTTCAAGATGGCATTCGTTGAATTGGTAA
 CATAGTA ACTATCCACTCCCTTCAGTTTAGCTGCCTCTTGAACCCAGGATTCCTTGCGGTT
 TTGGCGGTTCAACAGGAATCTTTTTCTTTTCCAGAAACCGTAAAAGCTGATTGTTTCTG
 AGTAAAAGACCCATCTTTACTTTTTTTAGGAGAGAAAAAGACGCTAATATTTTTCTGAGA
 TTTAGTCATATCTTTATTGACTTGACGAGATAGGGAATCACCCAAAGCCATAATCACAAC
 AACTGATGAAACACCGATAATAATCCCAATCATAGTAAGCAAAGAACGCATCTTGTGAGC
 CATGATAGATGAAAAGGCAAATTTTCAGATTCTGCATCTTAGTTTTCTCTCTTCTTA ACT
 GAGCACTGTCAGACGAAATGACCCCATCCCGAATGACAATCTGACGTTTGGCATAGGCAG

CAATCTCAGGCTTCATGCGTTACCATGATAATGGTTTTTCCTTCTTTATTCAAATCAACC
AATAATTGCATAAATTTGGTTACCTGTTTTTGGTATCCAAGGCTCCTGTCGGTTCATCCGCT
AGGATAATAGAAGGATTGTTTACCAAGGCACGCGCAATGGCTACACGTTGCTTTTGACCA
CCAGATAATTCTGAAGGTAAATGGTGACTACGTTCTATCAATTCAACCTTGTCTAAATAT
TCCTCAGCCAACTTGCGACGTTTTGAAGACGAAACTCCTGCGTAAATCAAGGGCAATTCT
ACATTTTGCAGAGCATTGAGCTTCGATAGAAGAAAGAACTGCTGAAAGACAAAACCGATT
TGTTGGTTACGGACCTTAGCTAGTTGTTTTTACCAAGCCCAGCCACTTCTTGACCTTCA
AGATAATATTCTCCACTGGTTGGTGTATCCAACATGCCAATCGTATTTCATCAGAGTGGAC
TTACCAGACCCAGATGGTCCCATGATGGCTACAAATTCACCCTCATTCACTTCTAGATTG
ATATTTTTTGAGAACCTGCAGTTCCTGGTCAACATTACGGTAACTTCTGAAGATATTTTTT
AGACTAATTAGTTGCTTCATCAGCCTTCACCTCTTTTCTTCCAAGGAAGATGTTGG
ATTACTGATGACCTTAGCACCGTTTCGTTAAACCAGAAGTGATTTCTTGATTTTCTGCGTC
AGCATTTCCCAATGAAACCTCAACTTTTTTAGCCTTTTGTTGTTTCATCCACAATCCAGAC
ATAATTTTTACTATCATCCATTACTAGACTGCTAACAGGAACAAGAATAGCCTTAGTTTTT
GCTTTTAACCTCAATGTTGACAGAAAAACCTTGTTTCAAATCACCAACCTCGCCTGTCAC
ATCAATAGTATAAGGGTATTTAGAACCTGTATTATTCCCGGCTGCTGGACTAGCTGCTTC
ACCATTGTTTTTTAGGATAGTCAGAAATATAGGCTTAATTTCCCAGTCCATTTTTTATCAG
GATACACTTTAGAAAGTAAAGCTTACTTCTTGACCTACAGAAAGGTTGGCTAGATTGTACT
CAGACAATTCTCCCTTGACTTGTAATTTTCATTGCTGACAATATGAACCATAACTTGAC
TCGCCCTGTTGGAGATTTAGAAACATTGCTATTGACTTCGACTACAGTTCCCTCTAGGG
TACTGAGAACAGTTGTTGCATCCAATTGACTTTGAGCCTTGCTTAATTGCGCTGCAGCAT
CTGCACGCGCATCACGGGCATCACCCAATTGAGCATCAATAGAAGCAACAGAATTTCCAG
CCACTGGAGTTGGGCTTTGCACCGTTGCATCTTCTCCTCCTACTGGCGCTGGTAACTGTG
GAGCCTGAGCTGAAGCGGCTTCATTTCTGCTTGATTGAGTTCATTGATATGACGATCTG
CCTTAGCTACTGCTCGACTAG

ORF Predictions:

ORF #	Start	End	Direction	Length
1	1084	1380	R	99 aa

> 3864194-3 ORF translation from 1084-1380, direction R
VTGEVGDLDKQGFVSNIIEVKSKTKAILVPVSSLVMDDSKNYVWIVDEQQKAKKVEVSLGNA
DAENQEITSGLTNGAKVISNPTSSLEEGKEVKADEATN*

Description:

unknown

Assembly ID: 3864338

Assembly Length: 1335bp

> 3864338 Strep Assembly -- Assembly id#3864338

ATCGAATTCCCTATTTTAACACTTTCTTTTCTAAACAGTCTATATTTTATTTCAAACCTG

TATTATATTTTTGAAAAATAAAAGTCCTTTTTTCTTTTTTTCAGAAAAAGGGTATAATA
 AAAGAAAATAAGCAGTAACACTCAATGGAAATCGAAAAAGCAAAC TAGGAAGCTAGCCGC
 AGATTGCTCAAAACACTGTTTTGAGGTTCGAGATAGAGCTGACGTGGTTTTGAAGAGATTT
 TCGAAGAGTATAAAAAGGTGCTAGGCATGTTGATTTTTCTTTTGTTAAATGATTTGTCAA
 GAAAAATCATCCATATTGGACATGGATGCCTTTTTTGCTGCAGTGGAAATCAGGGATAAT
 CCTAAACTCAGAGGAAAACCTGTCATTATTGGAAGCGACCCTCGGCAAACAGGTGGACGG
 GGAGTCGTTTTCTACCTGTAGTTATGAGGCAAGAGCTTTTGGTGTCCATTCTGCCATGAGT
 TCCAAGGAAGCTTATGAACGTTGTCCCCAGGCTGTCTTTATCTCAGGGAATTCGATGAGA
 AATACAAGTCTGTGGGACTCCAGATTCGAGCTATTTTTAAGCGCTATACAGATTTGATTG
 AACCCATGAGCATTGACGAAGCCTATTTGGATGTGACAGAAAAATAAACTCGGTATCAAGT
 CAGCGGTCAA AATTGCTCGCCTCATTCAAAAAGATATCTGGCAAGAACTCCATCTAACTG
 CTTCCGCAGGCGTTTTCTTACAACAAATTTCTTAGCTAAAAATGGCGAGTGATTATCAAAAAC
 CACATGGTTTTGACAGTGATTCTACCTGAACAGGCTGAGGATTTTCTCAAACAAATGGATA
 TTTCCAAATTTTCATGGAGTAGGAAAAAAGACAGTAGAACGTCTTCATCAAATGGGCGTTT
 TTA CTGGTGCTGATTTACTTGAAGTTCCTGAGGTAACCCTAATAGACCGTTTTTGGTAGAC
 TAGGCTATGATCTGTATCGAAAGGCTCGTGGCATTCACTCACTCCAGTCAAATCCAATC
 ACATCCGTAAATCAATCGGCAAGGAGAAAACCTACGGGAAGATTCTCCGTGCTGAGGAAG
 ATATCAAAAAGAGAGCTGACTCTTCTATCAGAAAAAGTCGCTCTCAATCTACATCAACA
 AGAAAAAGCTGGAAAAATTGTCATTTTGAAAATCCGCTACGAGGACTTTTCAACTCTTAC
 CAAACGAAAAAGTATTGCTCAAAAACACAAGATGCTAGTCAGATAAGCCAAATAGCCCT
 GCAACTCTATGAAGAATTAAGTGAGAAAGAAAGAGGTGTCCGCCTATTGGGGATTACCAT
 GACTGGATTTTAAAG

ORF Predictions:

ORF #	Start	End	Direction	Length
1	552	1100	F	183 aa

> 3864338-2 ORF translation from 552-1100, direction F
 VGLQIRAIKRYTDLIEPMSIDEAYLDVTENKLGKISAVKILRIQKDIWQELHLTASAG
 VSYNKFLAKMASDYQKPHGLTVILPEQAEDFLKQMDISKFHGVGKKTVERLHQMGVFTGA
 DLLEVPEVTLIDRFGRLLGYDLYRKARGIHNSPVKSNHIRKSIGKEKTYGKILRAEEDIKK
 ES*

Description:

ECODINJ NCBI - Escherichia coli (sub_strain W3110, strain K-12)
 DinP, DNA damage inducible protein

Assembly ID: 3864360

Assembly Length: 1796bp

> 3864360 Strep Assembly -- Assembly id#3864360

TCCAAGCTAGCTATTTCTGTGGAAGGGGCTTCGGTTGGCAGAACCTGGTGAATTTACCCAA

ACGTGCTTTTTTAAACGGTCGCGTAGACTTGACACAGGCAGAGGCTGTGATGGATATCAT
 CCGTGCCAAGACTGACAAGGCCATGAACATTGCGGTCAAACAATTAGACGGCTCCCTTTC
 TGACCTCATTAACAATACCCGTCAAGAAATCCTCAATACACTTGCCCAAGTTGAGGTCAA
 TATCGACTATCCTGAATATGATGATGTTGAGGAAGCTACTACTGCCGTTGTCCGTGAGAA
 GACTATGGAGTTTGAGCAATTGCTAACCAAGCTCCTTAGGACAGCACGTCGTGGTAAAAT
 CCTTCGTGAAGGAATTTCAACGGCTATCATTTGGACGTCCCAACGTTGGGAAATCAAGCCT
 TCTCAACAACCTCTTGCGTGAGGACAAGGCTATCGTAACCGATATCGCTGGGACAACACG
 AGATGTCATCGAAGAGTACGTCAACATCAATGGTGTTCCTCTAAAATTGATTGACACAGC
 TGGTATTTCGTGAAACGGATGATATCGTTGAACAAATCGGTGTTGAGCGTTCGAAAAAAGC
 CCTCAAGGAAGCCGACTTGGTTCTACTAGTGCTAAATGCCAGTGAACCACTGACTGCGCA
 AGACAGACAACCTTCTTGAAATTAGCCAAGATACCAATCGCATTATTCTACTTAATAAAAC
 CGACCTGCCAGAAACGATTGAAACTTCGAAACTACCTGAAGACGTTATCCGTATTTTCAGT
 CCTTAAAAACCAAAACATCGACAAGATTGAAGAGCGAATCAACAACCTCTTCTTTGAAAA
 TGCTGGCTTGGTCGAGCAAGATGCTACTTACTTGTCAAACGCCCGTCACATTTCCCTGAT
 TGAAAAAGCAGTTGAAAGCCTACAAGCCGTTAATCAAGGTCTTGAGCTGGGGATGCCAGT
 TGATTTGCTTCAAGTTGACTTGACTCGTACTTGGGAAATCCTCGGAGAAATCACTGGGGA
 TGCTGCTCCAGATGAACTCATCACCCAACCTCTTAGCCAATTCTGTTTAGGAAAATAAGA
 AAAATCCATGATCCTTCATTTCGGTCATGGATTTTATTGTCTTTATTAGTAATCTGGTCTT
 AAGACCCCTGTTACAGTTGCCTTAGTTGCTTCGTAGTCGCCATCTACGACAACCTTGATA
 ATGCGTTTGACATCTTCTTCTGGTGCTGGAACAAGAGGTAGACGAGTGGGTCCAGCTTCA
 AATCCCATATAGTTAAGAATTGCCTTAAGTGGAGCAGGACTTGGATAAGAGAAGAGAGCA
 TTAACCTTAGGAATGAATTTACGCTGAATTGCTGCGGCTTTCTTCATATCGCTTTCTGCA
 ATGGCAGTAAACATCTCGTGCAATTCATCCCCATTTGTATGAGAGGCAACAGAAATAACC
 CCATCCGCCCCAAGGTTTCATGGCATGGAAGCATCTCCATCCTCACCTGTATAAATCAAG
 AACTCTTCAGGCTTGTGCTCAATCAAGTAAGCCATATTAGCCAAGCTAGTACATTCTTTG
 ACACCGATAATATTTGGATGGTCAGCCAAGCGAAGCATGGTTTCTGGAGTCAATTGACA
 ACTACACGCCCTGGAATGTTATAGATAAATAATTGGTAGGTCAGAAGCATCTGCAATAGCC
 TTAAGTGCTGATACATCCCTTCTTGAGAAGGTTTGTGTAGTAAGGAACAATAGCAAGC
 CCAGCTGCGAAACCACCAAATTCGCTACTTCTTTGACAACTCAATAGAGTCACG

ORF Predictions:

ORF #	Start	End	Direction	Length
1	47	1078	F	344 aa

> 3864360-1 ORF translation from 47-1078, direction F
 VNLPKRAFLNGRVDLTQAEAVMDIIRAKTDKAMNIAVKQLDGSLSDLINNTRQEILNTLA
 QVEVNIDYPEYDDVEEATTAVVREKTMEFEQLLTKLLRTARRGKILREGISTAIIGRPNV
 GKSSLLNNLLREDKAIIVTDIAGTTTRDVIEEYVNINGVPLKLIDTAGIRETDDIVEQIGVE
 RSKKALKEADLVLLVLNASEPLTAQDRQLEISQDTNRIILLNKTDLPETIETSKLPEDV
 IRISVLKNQNIIDKIEERINNLFENAGLVEQDATYLSNARHISLIEKAVESLQAVNQGLE
 LGMPVDLLQVDLRTWEILGEITGDAAPDELITQLFSQFCLGK*

Description:

THIOPHENE AND FURAN OXIDATION PROTEIN THDF. - ESCHERICHIA COLI.

Assembly ID: 3864388

Assembly Length: 2337bp

> 3864388 Strep Assembly -- Assembly id#3864388

CTTCGTACAGGTGGTTCCTATGCAAGGGTGGAAAGCCAATCGTCAGAACAACAAGCATCTT
CATCAAGCCAGAACTGGAGCAATTACAAAAAGAAATTGCTGAAGAAGAAGCAAGCTTGGG
TTCAGAAGAAGTGGCTTTGAAGACCTTGCAAGATGAGATGGCCAGATTGACCGAGTCATT
AGAAGCTATTAAATCTCAAGGAGAGCAGGCACGTATTCAGGAGCAAGGCTTGTCCCTCGC
TTATCAGCAAACCTAGTCAGCAAGTTGAAGAACTGGAAACTCTTTGGAAACTCCAAGAAGA
GGAAATAGATCGTCTTTCCGAGGGAGATTGGCAAGCGGATAAGGAAAAATGCCAAGAGCG
TCTTGCTGCAATCGCCAGTGACAAGCAAAATCTGGAAGCTGAGATTGAAGAGATTAAGTC
TAATAAAAATGCCATCCAAGAACGCTATCAAACTTGCAGGAAGAGCTAGCGCAAGCTCG
TTTGCTTAAGACAGAACTGCAAGGGCAAAAACGTTATGAAATTGCTGATATTGAACGCTT
AGGCAAGGAATTGGACAATCTTGATTTTGAACAAGAGGAAATCCAGCGCCTTCTTCAAGA
AAAGGTTGACAATCTTGAGAAGGTTGATACAGAATTGCTCAGTCAACAGGCGGAAGAATC
CAAACTCAGAAAACGAACCTCCAACAAGGTTTGATTGCGCAAACAGTTTGAGTTGGATGA
TATAGAAGGTCAGCTGGATGATATTGCTAGTCATTTGGATCAGGCTCGCCAGCAGAATGA
GGAGTGGATTGCGCAAGCAAACACGTGCTGAAGCTAAGAAAGAAAAGGTCAGCGAGCGCTT
TGCCGCCATCTACAAAGTCAATTAACAGACCAGTACCAGATTAGCCATACTGAAGCTCTA
GAAAAAGCGCATGAATTGGAAAACCTCAATCTGGCAGAGCAAGAAGTTAAGGATTTAGAG
AAGGCTATTTCGCTCACTGGGTCCCTGTCAATATAGAAGCTATTGACCGGTACGAAGAAGTT
CACAACCGTCTGGACTTTCTAAATAGTCAGCGAGATGATATTTTGTGTCAGCGAAAAATCTG
CTCCTTGAAACCATTACAAAGATGAATGATGAGGTTAAGGAACGCTTTAAATCAACCTTT
GAAGCTATTTCGTGAGTCCTTTAAAGTGACCTTCAAGCAGATGTTTGGCGGAGGTCAGGCA
GACTTGATATTGACTGAGGGCGACCTTTTACAGCTGGTGTGGAGATTTCTGTTCAACCTC
CAGGTAAGAAAATCCAGTCGCTTAACCTCATGAGTGGTGGTGAAAAAGCCCTATCGGCTC
TTGCCTTGCTTTTCTCCATTATTCGTGTCAAGACCATTCCTTTTGTGTCATCTTGGATGAGG
TGGAAGCTGCGTTGGATGAAGCCAATGTTAAACGTTTTGGGGATTACCTCAACCGCTTTG
ACAAGGACAGCCAGTTTATCGTCGTAACCCACCGTAAGGGAACCATGGCAGCGGCCGATT
CCATCTATGGAGTGACCATGCAAGAATCGGGTGTTTCAAAGATTGTTTCAGTTAAGTTAA
AAGATTTAGAAAGTATTGAAGGATGACAATTAACTAGTAGCAACGGATATGGACGGAAC
CTTCCTAGATGAGAATGGGCGCTTTGATATGGACCGCCTCAAGTCTCTCTTGGTTTCCTA
CAAGGAAAAAGGATTTACTTTGCGGTGGCTTCGGGTCGGGGATTTCTGTCTCTGGAAAT
CGAATTATTTGCTGGTGTTCGTGATGACATTATTTTCATCGCGGAAAATGGCAGTTTGGT
AGAGTATCAAGGTCAGGACTTGTATGAAGCGACTATGTCTCGTGACTTTTATCTGGCAAC
TTTTGAAAAGCTGAAAACGTCACCTTATATAGATATCAATAAACTGCTCTTGACGGGTAA
GAAGGGTTCATATGTTCTAGATACGGTTGATGAGACCTATTTGAAAGTGAGTCAGCATT
TAATGAAAATATCCAAAAAGTAGCGAGTTTGGGAAGATATCACAGATGACATTTTCAAAT
TACAACCAACTTCACAGAAGAAACGCTAGAAGCTGGTGAAGCTTGGGTCAATGATAATGT
CCCTGGTGTCAAGGCTATGACAACCTGGCTTTGAATCTATTGATATTGTTCTGGACTATGT

CGATAAGGGTGTAGCTATTGTTGAATTAGCTAAAAAACTTGGCATCACAAATGGATCAGGT
 CATGGCTTTTGGAGACAATCTTAATGACTTACATATGATGCAGGTTGTGGGACATCCTGT
 AGCTCCTGAAAATGCACGACCAGAGATTTTAGAATTAGCATAAGACTGTGATTGGTC

ORF Predictions:

ORF #	Start	End	Direction	Length
1	1239	1586	F	116 aa

> 3864388-3 ORF translation from 1239-1586, direction F
 VEISVQPPGKKIQSLNLMMSGGEKALSALALLFSIIRVKTIPFVILDEVEAALDEANVKRF
 GDYLNRFDKDSQFIVVTHRKGTMAAADSIYGVTMQESGVSKIIVSVKLDLESIEG*

Description:

P115 protein - *Mycoplasma hyorhinis* (SGC3) (similarity to
 SMC1_YEAST, chromosome segregation protein)

Assembly ID: 3864406

Assembly Length: 2162bp

> 3864406 Strep Assembly -- Assembly id#3864406

CTAAAAGTGAAGCCCGATAGCGTCTCTCTCCTGCAAGGATTTTCATAACCAATAACAGGAG
 ATTGACGAACAATAATCGGTTGAATGACCCCATTTTCTTTGATAGACTGTGCTAGTTCAT
 CTAGCTTTTCTCTATCAAATTCTTTTCGGGGTTGATAGGGATTTTTTTGTATATCTGTGA
 TAGAAATCATTTCAAATTTTCCATGATTCTACACTAACACATCTTTTCTCTTATGTAAA
 GCTTTCTTTACATAGATGTCAATTAAGATTCTAAATCACCTGAACTCTTGTTAAGTTTGA
 TAGAGGTAGTTTCTTCTTTCCCGTTACGATAGTAGGTTATCTTAATGGTGTCTCCGATAG
 AATGGTTGTAAAGAGCACTTTGTAAAGTCTGTTGATGAAGCAATCTCTTTGTCATCTACTT
 TTGTAATTACATCGTATTTTTCAAGGTGACCATTGGCAGGCATATTACTTTGTACCGAAC
 GAACAATTACACCAGATGTAACATTACTTGGAATATTGAGTCTTCTGATGTCGCTTGTA
 TCACATTAGATAAATTAACCATCTGGATTCCCAAAGCTGGACGCGTCACCTTTCCGTTTTT
 TTTCTAACTGTTCAATAATATTGATAGCATCATTTGCAGGAATTGCGAAACCAAGACCTT
 CTACAGATGTTCCCTCCATTTGTAGCAATTTTACTTGAGGTAATTCCGATAACCTGCCCTT
 GAATATTGATCAGTGGGCCGCCAGAGTTACCTGGGTAAATAGCAGTATCAGTTTGATGG
 CTTTTGTAGAAATAGCTTGTCCATCTTCCGATTTTAAAGGATACATTTCTATTGAGACTGG
 ATACGATACCTTGAGTGACAGTATTTGCATATTCAGAACCTAACGGGCTACCGATGGCAA
 TAGCAGTTTCTCCTACAGTTAACTTACTAGAATCACCAAACCTCAGCTACTGTTGTCACTT
 TTTCTGAAGAGATTTTCGACGACAGCAATATCAGAGAAAGTGTGAGCTCCGACAATTTCTC
 CAGGTACTTTAGTCCCCTCTGACAATCGAATATCTACTTTGCTGGCGCCATTTATAACGT
 GATTGTTGGTGACGATGTAAGCTTCTTTATCATCTTTTTATAAATAACTCCAGATCCTT
 CACTAGAGATTCGCTGAGAATCTGTGTGAGTATCATCATTTGCCAAATACGCTATTTTGTC
 TGTTTGCCGAATAAGTAATAACAGAAACAACAGCATCTTTTACTTTGTTAACGGCCTGTG
 TTGTTGAATTTTCCGTTCTTATAGGCAGTTTGTGTAATAGTACTATTGTTGTTAGAGTT

GTTTACACTACTTTTTTTGAGTTAGTTGAGTTATTGAAAACTACCCAAGGCTCCACTAAA
 AAAGCTAATGACGATAACGACTAATAATTGAAACCATTTTTTGTAAAATGTTTTTAGATG
 TTTCATATTTGCCTCCATATGTTTGAATTACTGAAAGTATAAACTGACTAGCTTAATTAT
 AACTTAAACACAAAAGTTTTACACAACTGTGGATAACTCTTTTGAACTGTGATTTTCT
 TAATTGAAATCTATTTTTTATTTTGTGAATAAGATGTGAAAAATAGAGAATATGTTAGA
 ATAGAGTCATGAAAATTAAAGTTGTAACAGTTGGGAACTGAAAGAAAAGTATTTAAAG
 ATGGTATCGCAGAGTATTCAAAACGAATTTCTAGATTTGCTAAGTTTGAAATGATTGAGT
 TATCAGATGAAAAACACCAGATAAGGCCAGTGAATCAGAAAATCAAAAGATTTTAGAAA
 TAGAAGGTCAGAGAATTTTATCAAAAATTGCTGACCGTGATTTTCGTTATTGTGTTAGCCA
 TTGAAGGGAAAACCTTTCTTCTCAGAAGAATTTAGTAAGCAGTGAGAAGAACTTCTATAA
 GGAAGGATGTCTACTCTTACTTTTATTATTGGGGGAAGTTTAGGATTGTCATCATCTGTA
 AAAAATAGAGCCAATCTTCTGTCAGTTTGGTTCGCCTAACCTTGCCTCATCAGTTAATG
 AGACTAGTTCTTGTGTAACAAATCTATCGCGCTTTTACGATTCAGCAGGGATTCCCCTAC
 CATAAATAGAGAATTGACTTTTAATTGAATTTTGGTAGAATAATTGTGTTAGGTCTCAT
 AG

ORF Predictions:

ORF #	Start	End	Direction	Length
1	263	958	R	232 aa

> 3864406-1 ORF translation from 263-958, direction R
 VTTVAEFGDSSKLTVGETAIAIGSPLGSEYANTVTQGI VSSLNRNVSLKSEDGQAISTKA
 IQTDTAINPGNSGGPLINIQQQVIGITSSKIATNGGTSVEGLGFAIPANDAINIIEQLEK
 NGKVTRPALGIQMVNLSNVSTSDIRRLNIPSNVTSGVIVRSVQSNMPANGHLEKYDVITK
 VDDKEIASSTDLQSA LYNHSIGDTIKITYYRNGKEETTSIKL NKSSGDLES*

Description:

Bacillus subtilis (strain 168,) DNA. Homologous to E. coli
 serine protease HtrA (BLAST)

Assembly ID: 3864452

Assembly Length: 1766bp

> 3864452 Strep Assembly -- Assembly id#3864452

ATCGAATTTTCCAAAATGGGGAGCTAGAGCAGTGGAGTGATTATGTGGCAGACGATTTGA
 TTCAGCATAATCATGAGATTGGACAAGGAAGTGCTGCTTATAAAAACCTATGTGGCTGAAT
 ATATTGTCACTTTTGACTTCGTTTTTCCAACCTCTTAGGACAAGGAACTATGTGGTTAGCT
 ATGGTCAGACTCAGATTGATGGCGTTGCTTATGCCAAGTACGATATCTTCCGTTTAAAGA
 ACGGGAAAATTGTGGAGCATTGGGATAATAAGGAAGTCATGCCTAAGGTAGAAGACTTGA
 CCAATCGAGGGGAAGTTTTAAATTGAGGACAAAGAATGATTGAATACAAAAATGTAGCACT
 GCGCTACACAGAAAAGGATGTCTTGAGAGATGTCAACTTACAGATTGAGGATGGGGAAATT
 TATGGTTTTAGTAGGGCCTTCTGGGTCAGGTAAGACGACCATGCTCAAGATGATTAACCG

TCTTTTGGAAACCAACTGATGGAAATATTTATATGGATGGGAAGCGCATCAAAGACTATGA
 TGAGCGTGAACCTTCGTCTTTCTACTGGTTATGTTTTACAGGCTATTGCTCTTTTTTCCAAA
 TCTAACAGTTGCGGAAAATATTGCTCTCATTCCTGAAATGAAGGGGTGGAGCAAGGAAGA
 AATTACGAAGAAAACAGAAGAGCTTTTGGCTAAGGTTGGTTTACCAGTAGCCGAGTATGG
 GCATCGCTTACCTAGTGAATTATCTGGTGGAGAACAGCAACGGGTTCGGTATTGTCCGAGC
 TATGATTGGTCAGCCCAAGATTTTCTCATGGATGAACCCTTTTCGGCCTTGGATGCTAT
 TTCGAGAAAACAGTTGCAGGTTCTGACAAAAGAATTGCATAAAGAGTTTGGGATGACAAC
 GATTTTTTGTAACCCATGATACGGATGAAGCCTTGAAGTTGGCGGACCGTATTGCTGTCTT
 GCAGGATGGAGAAATTCGCCAGGTAGCGAATCCCGAGACAATTTTAAAAGTGCCTGCAAC
 AGACTTTGTAGCAGACTTGTTTGGAGGTAGTGTTCATGACTAATTTAATTGCAACTTTTC
 AGGATCGTTTTAGTGATTGGTTGACAGCTACAATGACATTGGTCGGTTCCTTGAGCAAGA
 GATAGATTAGCCAGACAGTCATGCCCAAATCCCTCCAGGTAAGAGCATAGACCGTTGCA
 CATTAAGTACGATTAAAAAAGTGATAATGGCAAGAAAACCTTGCTACTGCTTGTAAATAAAA
 AGGTTGTTAGTGTCATATTAGTTTCATCAATACCAAGGCGACAGAAGTTCCTGCCCTAAA
 GCGAGGGTAATGAGCAGGGATTCAAACATCTTACTCATACCAGAGTTTATGTGGTTGGTC
 ATAATATCACGGACCGCATTGGTCAAGGCAATACCTGGTACAAACGGCATGACCGCACCA
 GCTATAATCAAATCTGCCGTTGAAGGAAAACCTGTGTAGCGAGCCCAAACTGGGCAATT
 ATCCCAAAGACAAAAGCTCCAGCAAAGGCTGTCAAAAGGGAATTCGGATAAATTTTCC
 ACATAGAGGGAAAAGGCCAAAACCAAATAAGGTCCGCACTCCTGCCCCAAGTGCGTCGTAG
 ATATTTCCGCTAAACATAACTGAAAAGAAAGGAGCACTAAAGGTTCGAGCCAGAGTTACC
 TGCAACTTAGTATAGGGAAGGGGTTGAGCTTGCAAGGCCGTCAATTGCTTAAAGGCTGTT
 TCTAAGTCAATCTGCCCCCAACTGG

ORF Predictions:

ORF #	Start	End	Direction	Length
1	1079	1201	R	41 aa

> 3864452-2 ORF translation from 1079-1201, direction R
 VQRSM LLPGGILGMTVWLIYLLLKEPTNVIVAVNQSLKRS*

Description:

unknown

Assembly ID: 3864458

Assembly Length: 1705bp

> 3864458 Strep Assembly -- Assembly id#3864458

CTCTGACGGAGGCTGGTTATGTGGGTGAGGATGTGGAAAATATACTCCTCAAACCTCTTGC
 AGGTTGCTGACTTTAACATCGAACGTGCAGAGCGTGGCATTATCTATGTGGATGAAATTG
 ACAAGATTGCCAAGAAGAGTGAGAATGTGTCTATCACACGTGATGTTTCTGGTGAAGGGG
 TGCAACAAGCCCTTCTCAAGATTATTGAGGGAACCTGTTGCTAGCGTACCGCCTCAAGGTG
 GACGCAAACATCCACAACAAGAGATGATTCAAGTGATACAAAAAATATCCTCTTCATCG

TGGGTGGTGCTTTTGATGGTATTGAAGAAATTGTCAAACAACGTCTGGGTGAAAAAGTCA
TCGGATTTGGTCAAACAATAAGGCGATTGACGAAAACAGCTCATACATGCAAGAAATCA
TCGCTGAAGACATTCAAAAATTTGGTATTATCCCTGAGTTGATTGGACGCTTGCCTGTTT
TTGCGGCTCTTGAGCAATTGACCGTTGATGACTTGGTTCGCATCTTGAAAGAGCCAAGAA
ATGCCTTGGTGAAACAATACCAAACCTTGCTTTCTTATGATGATGTTGAGTTGGAATTTG
ACGACGAAGCCCTTCAAGAGATTGCTAATAAAGCAATCGAACGGAAGACAGGGGCGCGTG
GACTTCGCTCCATCATCGAAGAAACCATGCTAGATGTTATGTTTGAGGTGCCGAGTCAGG
AAAATGTGAAATTGGTTCGCATCACTAAAGAAACTGTGCGATGGAACGGATAAACCGATCC
TAGAAACAGCCTAGAGGTGACTATGGAACCTTAATACACACAATGCTGAAATCTTGCTCAG
TGCAGCTAATAAGTCCCCTATCCGCAGGATGAACTGCCAGAGATTGCCCTAGCAGGGCG
TTCAAATGTTGGTAAATCCAGCTTTATCAACACTATGTTGAACCGTAAGAATCTCGCTCG
TACATCAGGAAAACCTGGTAAAACCCAGCTCCTGAACTTTTTTAACATTGATGACAAGAT
GCGCTTTGTGGATGTGCCTGGTTATGGCTATGCTCGTGTTTCTAAAAAGGAACGTGAAAA
GTGGGGGTGCATGATTGAGGAGTAATTTAACGACTCGGGAAAATCTCCGTGCGGTTGTCA
GTCTAGTTGACCTTCGTCATGACCCGTCAGCAGATGATGTGCAGATGTACGAATTTCTCA
AGTATTATGAGATTCCAGTCATCATTTGTGGCGACCAAGGCGGACAAGATTCCTCGTGGTA
AATGGAACAAGCATGAATCAGCAATCAAAAAGAAATTAAACTTTGACCCAAGTGACGATT
TCATCCTCTTTTCATCTGTCAGCAAGGCAGGGATGGATGAGGCTTGGGATGCAATCTTAG
AAAAATTGTGAGGAAAAGAAAATGGCAAAAACAATTCATACAGATAAGGCCCCAAAGGCT
ATCGGGCCCTATGTTCAAGGAAAAATCGTTGGCAACCTTTTGTGTTGCTAGCGGTCAAGTT
CCCCATCCCCTGAAACTGGGGAAATTGTAGGAGAGAATATCCAAGAACAGACAGAGCAA
GTCTTGAAAAACATCGGTGCTATTTTGGCAGAAGCAGGAACAGACTTTGACCATGTTGTC
AAAACAACCTTGTTTCTTGAGCGATATGAACGACTTTGTTCCTTTTAATGAGGTTTACCAA
ACGGCCTTCAAAGAGGAATTCCCAG

ORF Predictions:

ORF #	Start	End	Direction	Length
1	797	1105	F	103 aa
2	1179	1391	F	71 aa

> 3864458-2 ORF translation from 797-1105, direction F
VTMELNTHNAEILLSAANKSHYPQDELPEIALAGRSNVGKSSFINTMLNRKNLARTSGKP
GKTQLLNFFNIDDKMRFVDVPGYGYARVSKKEREKWGCMIIE*

Description:

unknown

> 3864458-3 ORF translation from 1179-1391, direction F
VQMYEFLKYYEIPVIVATKADKIPRGKWNKHESAIKKKLNFDPSSDFILFSSVSKAGMD
EAWDAILEKL*

Description:

HYPOTHETICAL 22.0 KD PROTEIN IN LON-HEMA INTERGENIC REGION
(ORFX). - BACILLUS S UBUTILIS.

Assembly ID: 3864474
Assembly Length: 1673bp

```
> 3864474 Strep Assembly -- Assembly id#3864474
ACGTTTTGGGAAGTGTTCGGATAGCAGATTCCGAACAACTGATAATGGTTGGCAAAATC
ATTATTCCTAATAGTAACGAAGCTGGTTAGGACAACATCATGCCATTTCTTAAAAAGGTTT
TAATCCAAGGCACCAATAATTGTAGGCCGAAAAAACCATAAACAATAGATGGAATGGCTG
CCATCAAGTTGATAGCTGATTTTAAAGAAGCTATAGACGGGCTTTGGACAATTATAAACCA
TAAACACCGATGTCAAGATCGCCTGTTGGCACCCCAATCACAATCGCTCCTAAGGTCGAA
TAAATAAGGAACCAACGATCATTGGTAAAATACCATAGCTTGCCGGAATGTTTCGTTGGCG
ACCAATCACTGCCTAATAAAAAACGGGCAAAGCCGTAGTTAGCTATGAAAGGTAAGCCAT
TACTAAAAATAAAGAAACAGATTAGCAAAATAGCTACAACAGCTACTGTTGCACTCATGA
AAAAAATTGCCCTAAAAACTGCTTCTTTGAAGGCTTGTTTTGTCACATCTTGTCTTTCT
AGTGAAGAAAGTAAGGGAGATACGACACCTCCCTACTTGCCTTCTTTATCTTATTGTACG
ATGAAACGTCTGCATCTCTTTAGAGATTTATGGAGCAAACATTTTATTTAATCTTGTCCC
AGGTGGTTAATTTGCCACTAAAAACGTCCGCAAGTTCAGCCATACTGACTTGGCTTGCCT
TATTGTCATTATTGACCACAACAGCAATACCGTCTAAAGCAATAGCATCATGGGTGAGAC
TCTTACCTTCTTCAGGAGTTAATTCCTTAGAAACCATAACCAATATCAGCGGTTTTCTCCT
TAACAGCGGTAATACCTGCTGAAGACCCATTAGAGGTAATATCAATCGTAACTTCTGGAT
TTTCTTTTTTTATAAGCTTCTGCTAATTTTTCCATTAAAGAAGATACTGAAGTGGAACCTA
CAACAGACAACCTTGCCTGATAAGTGTGGCTTGTATATTCTGTGGTTTCGGTTTTTAGCTT
CAATAAATTTATTATCTGTGACCACTTGTGACCTTGTGTTGGAGTGGATAAAGCTGATAA
AATCTTGACCTAGCTTGGAAAGATTAGAAGACCAAACAATGTTGAAGGGACGTTGAAGAG
GGTATTCACCATCTAAACTGTGTCTCGACTAGCCTTGACACCATCAATCTCTAAAGCCT
TGACAGATTTTCGTTAAAGATCCCAAGGAGATGTAGCCGATAGCATTAGCATTCCTTGAA
CTGCTGAGAGAACACCTTCTGTACTATTTTGAATCACAGCTGTTTTTGGCAGTGTAGTCAA
TTTTTTTTATCACCGTCTTTTTTGAAGAATCCCTGTGATTTCTGTGAAGGCACCCCGTGTTC
CAGAGCCATTTTCTCGTGAAATCACCTCAATCGTTCCTGGAGCTGACTGTTTGGGAAGCAG
CTGACTGATTGCCACAGGCAACAAGCCCAAATCCTGATAAGCCAATGGCTGCAAGAGTAA
GCATTTTTTTTGAATTTTCATAATAATCACCTTTATCTCTATGTATTTTTCTTGTGTAGGCT
TACTACATTTATAGTCTAACAAGTCTTTGTAAAGGTTTATCCCTGATTCATGTAAAGATT
GTGTAAAGAATCAAAAAAAGCCACTTTTGAAAAATGGCTGCCCCCTAAAAATAG
```

ORF Predictions:

ORF #	Start	End	Direction	Length
1	68	247	R	60 aa
2	644	1528	R	295 aa

> 3864474-1 ORF translation from 68-247, direction R

VFMVYNCPKPVYSFLKSAINLMAAIPSIYVGFGLQLLPWIKTFLGNGMSCPNQLRYY*

Description:

PROBABLE ABC TRANSPORTER .PERMEASE PROTEIN (ORF72) . --BACILLUS SUBTILIS. (BLAST)

> 3864474-2 ORF translation from 644-1528, direction R
VIIMKFKKMLTLAAIGLSGFLVACGNQSAASKQSAPGTIEVISRENGSGTRGAFTEITG
ILKKDGDKKIDYTAKTAVIQNSTEGVLSAVQGNANAIGYISLGSLTKSVKALEIDGVKAS
RDTVLDGEYPLQRPFNIVWSSNLSKLGQDFISFIHSKQGGQVVDNKFIEAKTETTEYTS
QHLSGKLSVVGSTSVSSLMEKLAAYKKENPEVTIDITSNGSSAGITAVKEKTADIGMVS
RELTPEEGKSLTHDAIALDGIADVNNNDNKASQVSMAELADVFSGLTWTWDKIK*

Description:

probable hemolysin precursor - Streptococcus agalactiae (strain 74-360)

Assembly ID: 3864510

Assembly Length: 1702bp

> 3864510 Strep Assembly -- Assembly id#3864510
CTTTTTTATTTACACAAGTTCATAACGTGTCTTACTGGTGAAGGTTTGACCAGCTTTA
AGAATGACTTGGCCTTTAAGGTCACTGTGAATGGCATCTGGTAAAGCTTGCCTTCAAGA
GCAATCCCATTGTGCTGTAGCATTGGCTGACCTCCTATGATGACACTTTCATCCACAAAG
TTTGCTGTGTAGACCACAAAGCAAGGAGCTTCTGTCTTGAAAAGCAGGAAGCGACCTGAA
TTTTGGTCATAAAGGAATCCAGCATTGTCATGGCCTGCAGGAAGGGCAAATGGATGATCC
AAACCTGATGCCAGCTGGATTTGCTCATCTTCTTCTGCAAAGATATCCTTCAACAAGGCA
CCATTGTAGATGTGTTTGACCACATCACGGTTGGCTTCTGGAGTTTTGGCAGGAACACCG
TCAGGAGCGATTGAGTAAATGCCCTCTGTGTTTAGTTGGAAGACATGACGGTCAATCGTC
TGCGTGAAATCACCAGACAAGTTGAAATAGCTGTGGTTGGTTGGATTGACCAGCGTATCC
TGATCGGTCGTTACCTTGTAGATCGAATTCATGGAGGCACCAGTTTCTTCCAAGTGATAA
CTGATCGCCAAATCTTGAGATTTCCAGGGAACCTCCTGTCCCATCTGTACGCTCTGTGT
AGAGAGTCAAGCCATGATCGCTTACTTCTTCAACTTCAAACAAGCTGGAATCCCAACCAG
TTGAACCACTGTGATTACAGTTGCTAGCATTATTAACCTCAAGGTCATAGGTCTTACCAT
TGAGCTCAAAGGTCGCACCTGCAATACGACCCGCTACAGGACCTACACTTGCTCCATGCT
TGGGACTATTGCCTACATAACTATCAAAGTCATCAAATCCCAAGATAACATTGGCAAAAT
TTCCAGCCTTGTGAGGTGCGACATAGCGCAAGATAGTCGCACCATAAGTCATAACCTCAA
GTTGGTAGCCACCGTCTGTCTCAAATCGATAGGCCAAGACATCCTCACCTCAACATTTT
CAAATACACGCTCTGTGTATGCTTTTATTCTGTTCTCCTTTTACTATTTCTCTCAAGCAA
ACAAACCATAGAAAGCGTACTGACAATCTATGGTTTATCTGATAATTTACAAATCCTCTT
GTCAAGAATTCATAAACACTGTCTTACTTTTGATATTCGTGAATTATGACACCTTGACT
ACACGGTTTACTGTACCTGTAGGAGACGGTGTATCTGGTTTATTTTCTACCTTGAGTGAA
GTCAATAGGGCAAAGAGTTGGGCATAAACGATGTAAGGGAAGACACGGTAAATATCATTC

AAGACACCGCCACAACCAAGGGCCACTTCTTTGACATTTTCAAGACCAAAAGCTTGATCA
 CTCAAAGCACAACACGACGAGCAATCTGGTCACCAGCAACTTCACGAACCAAGTCCAAG
 TCGTACTTACGAGTGTAGTCCGTCGTTGTACCAAAGACCAAAACAACCTGTATTGTCGTTG
 ATAAGAGATTTTGGACCGTGACGGAAGCCAACTGGGCTTTCATACATGGTCGCAACTTGA
 CCAGCAGTTAATTCCAAAATCTTGAGCTGAGCTTCATGAGCAAGTCCAAAGAAAGGACCA
 GCGCCTAGAATAGATGACACGGTTAAAGTCTAAATCAACGAGATCTTTGACATCTTCTGC
 CTTGTCTAAAACTTTACGGGCA

ORF Predictions:

ORF #	Start	End	Direction	Length
1	1164	1640	R	159 aa

> 3864510-3 ORF translation from 1164-1640, direction R
 VSSILGAGPFFGLAHEAQLKILELTAGQVATMYESPVGFRHGP KSLINDNTVVLVFGTTT
 DYTRKYDLDLVREVAGDQIARRVVLLSDQAFGLENVKEVALGCGGVLNDIYRVFPYIVYA
 QLFALLTSLKVENKPDTPSPTGTVNRVVQGVIIHEYQK*

Description:

AGAS PROTEIN. - ESCHERICHIA COLI. (Probable tagatose-6-phosphate
 ketose/aldose isomerase)

Assembly ID: 3864526

Assembly Length: 1940bp

> 3864526 Strep Assembly -- Assembly id#3864526
 TGCAGGATTTGATTTGGACGACTTTTATTATTACCAGATTCGCCTAGGAATAGAAAAAAG
 AGCCCAAGAGTTGGACTATGATATCTTGCGCTATTTTAATGACCACCCTTTTACCCTAAG
 CGAGGAAGTGATTGGGATTCTCTGCATCGGAAAGTTTAGTCGAGCTCAGATTTCTGCCTT
 TGAAGAATACCAAAAGCCTCTTGATTTCTAGACAGCGATACACTTTCCTGGGACATAC
 CTGTATTATCACGGATTTTACACTGCTATGAAACAGGTTGTCGATTATTTCCCTCAGTCA
 AGGAATGGACCGTATCGGGATTCTAACAGGCCTTGAAGAAACAACAGACCAAGAAGAAAT
 CATTACAGGACAAGCGTCTAGAAAACCTCAAAAACCTACAGTCAAGCGAGGGGAATCTATCA
 TGATGAACTGGTCTTTCAAGGAAGATTTACTGCCAGTCTGGCTATGACTTAATGAAGGA
 GGCCATTACAGAGCTTGGGAGACCAACTTCCGCCAGCATTTTTCGCAGCCAGCGATAGTTT
 AGCTATCGGTGCCCTCCGTGCCCTCCAAGAAGCTGGAATCAGCCTGCCAGATCGCGTCAG
 CCTCATTTTCCTTTAACGACACTAGTCTGACCAAACAGGTCTATCCTCCCCTCTCTAGTAT
 TACAGTTTATACTGAAGAAATGGGCCGAGCAGGTATGGATATTCTTAACAAGGAAGTCCT
 CCACGGTCGGAAAATCCCTAGCCTGACCATGCTGGGAACCAGACTGACATTAAGAGAAAG
 TACCCTAAATCAAGAATAGGATAACATAAAAAACGAATAGAGTTCTAAAACCTCCTATTCG
 TTTTATTATTCGATTACAATCATAGACTTAATGGTCTTACGTTTCATCCATATCTTTGTAGG
 CTTGGTCGATATCTTCCAGTTTATAACTTGAAGTAAAGACGCGACCTGGATTGATATCAC
 CATCAAGGACGGCTTTTAGTAAAAATTGCTTATCGTATGTTGTAGCAGAAGCTGCCCCAC

CTGCTACAGAGATATTTTGCATAAATGTCGAACCAAGAGCACGATTATTATAGTGTGGGA
 CTCCTACAAAGCCCATACGCCCTCCATTATGAAGAACACCTAGCGCCTGTTCTATAGCAG
 CCTCCGTACCAACACATTCAAGTGCTGCGTCTGCTCCTCCGCCGAGGATTTACAGCACCT
 TGGTAATTCCTTCTTGACCACGTTCTGCAACAACAGCTGTGCGACCTGACTCCATAGCCA
 TCTTTTGACGGTCTTCATGACGGCTCATAAGGATAATTTGTGATGCTCCACGCATCTTAG
 CCGCGATGACAGCACATTGACCAACAGCCCCATCACCGATAACAACAACCTTGTCCCCTT
 TTTGAACATTTGCAACACGCGCCGCATGATAGCCTGTCGGCATGACATCTGCAAGAGTCA
 AAAGGGACTTGAGCATCCCTTCTGTATAGTCAGAAGGTTGACCAGGGATTTTAACCAGCG
 CCCAGTTTGCATAGTGGAAGCGAATATATTCTGCCTGAAAATCACCCCCCAAATTATTGC
 CAATATGATTGTGCAAGAACCGTCAAATCCAGCAAGACAGGCATCACACTCACCACATC
 CATGTGTAAAAGGGACAATCACAAAATCACCTGGTTTCACCGTCGTAATGGCTTCCCCAG
 CTTCTTCAACAATCCCAATCGCTTCGTGTCCACTTATTTTTTGTGTCCAACCTTTCGTTTT
 CCNTGGATTACGGTACCTCCATAAATTTGAACCACAAACGCACGCACGAACCACACGAAT
 AATCACATCATCCGCTTCTATTATTTGCGGACGTTCAATGCTAGCAAGTCCAACCTGACC
 TGCCTTTGTATATACTGCTGATTTTCAATTTAAAATTTTCCTTCCTTATAAAGTTTAATTTT
 GAGATTTAAACGATTTAAAG

ORF Predictions:

ORF #	Start	End	Direction	Length
1	845	1660	R	272 aa

> 3864526-2 ORF translation from 845-1660, direction R
 VKPGDFVIVPFTTHGCGECDACLAGFDGSCDNHIGNNLGGDFQAEYIRFHYANWALVKIPG
 QPSDYTEGMLKSLTLADVMPTGYHAARVANVQKGDVVGIDGAVGQCAVIAAKMRGAS
 QIILMSRHEDRQKMAMESGATAVVAERGQEGITKVREILGGGADAALCEVGTEAAIEQAL
 GVLHNGGRMGFVGVPHYNNRNLGSTFMQNISVAGGAASATTYDKQFLKAVLDGDINPGR
 VFTSSYKLEDIDQAYKDMDERKTIKSMIVIE*

Description:

ALCOHOL DEHYDROGENASE (EC 1.1.1.1). - ALCALIGENES EUTROPHUS.

Assembly ID: 3864548

Assembly Length: 2051bp

> 3864548 Strep Assembly -- Assembly id#3864548

ATCGAATTTTCTAGCCAGGCTACAGTTTTGGCAAGTAAGGTTTCATCTCAGGCAGTCAA
 CTGGGTGAGTGCTTTTATTAGCGGAGCTTCTCAAGTGATTGTTGCCTTGATTATCGTTCC
 TTTTCATGCTCTTTTATCTCTTGCGTGATGGGAAAGGCTTGCGTAACTATTTGACCCAATT
 CATTCCAAGAAAATTGAAGGAACCTGTTGGACAAGTTCTATCAGATGTGAATCAACAGTT
 GTCCAACCTATGTTTCGAGGGCAAGTGACAGTGGCTATTATTGTAGCAGTAATGTTTATCAT
 CTTCTTCAAGATTATTGGTCTACGCTATGCGGTTACGCTGGGGGTTACTGCTGGTATTTT
 AAATCTGGTCCCTTATCTTGGTAGCTTTCTAGCCATGCTTCCTGCCCTAGTATTGGGTTT

GATTGCTGGTCCAGTCATGCTTTTGAAAGTAGTGATTGTCTTTATTGTAGAACAAACTAT
TGAAGGCCGTTTTGTCTCTCCATTGATTTTGGGAAGTCAATTAAACATCCACCCTATTAA
TGTTCTCTTTGTTTTGTTAACTTCAGGATCTATGTTTGGTATCTGGGGAGTTTTACTTGG
TATTCCGGTTTTATGCCTCTGCTAAGGTTGTCATTTTCAGCCATTTTCGAATGGTATAAGGT
AGTCAGTGGTCTATATGAATTAGAGGGTGAGGAAGTCAAGAGTGAACAATAGTCAACAGA
TGTTACAGGCTTTGGAGGAGCAAGATTTAACTAAGGCTGAGCATTATTTTCGCCAAAGCTT
TAGAAAATGATTCAAGTGATCTTCTGTATGAGTTGGCAACTTATCTTGAAGGGATTGGTT
TCTATCCTCAGGCCAAGGAAATTTACCTGAAAATTGTAGAAGAATTTCCAGAGGTTTCATC
TTAATCTAGCTGCAATGGCTAGCGAGGATGGTCAAATAGAAAAAGCCTTTAACTATCTTG
AGGAAATCCAAGCTGACAGTGACTGGTATGTCTCGCTCTTTGGCTCTGAAGGCAGACCTA
TACCAGCTGGAAGGTTTGACAGATGTGGCACGTGAGAAATTATTGGAGGCCTTGACCTAC
TCAAAGGATTCTCTCTTGATATTGGGTTTGGCAAAGTTGGATAGTGAGTTGGAAAATTAC
CAAGCGGCTATTCAAGCCTATGCCCAGTTAGATAATCGCTCGATTTATGAGCAAACGGGC
ATTTCCACCTATCAACGAATTGGCTTTGCCTATGCTCAGTTAGGGAAATTTGAAACGGCT
ACTGAGTTTTTTAGAAAAAGCCCTGGAGTTAGAATACGATGACTTAACAGCTTTTTGAGTTG
GCCAGTCTTTATTTTGATCAAGAAGAATATCAAAAAGCCACCCTCTACTTTAAGCAGCTT
GATACCATTTCTCCTGACTTTGAAGGCTATGAGTATGGGTACAGTCAGGCTTTACATAAG
GAACATCAAGTTCAAGAAGCCCTGCGTATCGCTAAGCAAGGATTAGAGAAAAATCCCTTT
GAAACTCGCCTCTTGCTAGCTGCTTCACAATTTTCTTATGAATTGCATGATGCTAGTGGT
GCAGAAAATTATCTCCTTACTGCAAAAGAAGACGCTGAGGATACAGAAGAAATCTTGCTT
CGTTTAGCCACTATTTATCTGGAGCAGGAGCGTTATGAGGATATTCTAGACTTGCAGAGT
GAGGAGCCAGAAAATCTTTTGACCAAGTGGATGATTGCTCGTTCCTTATCAAGAAATGGAC
GATTTGGATACTGCTTATGAGCATTATCAAGAGTTGACAGGAGATTTGAAGGACAATCCA
GAATTTCTGGAACACTATATCTATCTCTTGCGTGAATTGGGACATTTTGAAGAAGCAAAA
GTCCATGCTCACACTTACTTAAACTGGTTCAGATGATGTGCAAATGCAAGAACTGTTT
GAGAGATTGTAAGAATGTTTAAACATATAGAAGTGTAGTTTATCTCTTTTGATAGCTACG
GTCTTTATTTGTACATGGTAGAATCTTTTACAAAAATACTTGGTAATCTTGTTTATTCA
TGCCATAATAG

ORF Predictions:

ORF #	Start	End	Direction	Length
-----	-----	-----	-----	-----
1	687	1055	F	123 aa
2	979	1932	F	318 aa

> 3864548-2 ORF translation from 687-1055, direction F
VRKSRVNSQQMLQALEEQDLTKAEHYFAKALENDSSDLLYELATYLEGIGFYPPQAKEIY
LKIVEEFPEVHLNLAAMASEDGQIEKAFNYLEEIQADSDWYVSLFGSEGRPIPAGRFDRC
GT*

Description:

unknown

> 3864548-3 ORF translation from 979-1932, direction F —
 VTGMSRSLALKADLYQLEGLTDVAREKLLLEALTYSKDSLILGLAKLDSELENYQAAIQA
 YAQLDNRSIYEQTGISTYQRIGFAYAQLGKFETATEFLEKALELEYDDLTA FELASLYFD
 QEEYQKATLYFKQLDTISPDFEGYEYGYSQLHKEHQVQEALRIAKQGLEKNPFETRLLL
 AASQFSYELHDASGAENYLLTAKEDAEDTEEILLRLATIYLEQERYEDILDLOSEEPENL
 LTKWMIARSYQEMDDLDTAYEHYQELTGDLKDNPEFLEHYIYLLRELGHFEEAKVHAHTY
 LKLVPDDVQMQELFERL*

Description:

unknown

Assembly ID: 3864582

Assembly Length: 1318bp

> 3864582 Strep Assembly -- Assembly id#3864582

CTTTAGCAATCAGTTTATTGGGAGATTTGACTGCCACTTCTGTTGGAACCTTGATAATCT
 TTTTACCCTCAAAGCGTTCATACCAGAAATCTTAACATCAACTGCTAAAATAACTACAT
 CCGCTGCATCAATCTGCTCTTGACTCAATTCATTTTCTACCCCTATTGTCCCCTGAGTCT
 CAACATGAATCACATGTCCAGCTACCTTTGCGGCATTCTCTAATTTTTCTGTGCAATAT
 AAGTGTGGGCAATTCCCATAGTACAAGCTGCAACACCAACAATTTTCATACGGATACCCT
 CCAAATTTTTTCTTATTAACAAAAAGCTGCAATCACATCATCAGATGTCTGAGCCCGAA
 CTAATTTGGCAACAACTTCGTCATTACCAAGTTTTTCGAGCAAAGAGTGATAAGGTCTTCA
 AATGCTCCCTAGCAGCTTCTGTATCATCACCACCTGCAAAGAGTACAATTACTTTGACCC
 CTTTCCCATCAATGGTCTCCCAAGGAATCTCATTTGTGATTTATAGCTATGACTACCCCG
 CCTTCTCCACAGCAGAACTCTAGCTATGGGGAATAGCAATATAATCCCAATACCGGTCT
 GTCCTTCTGCCTCTCTCTGATAAAGACCTTCGATAAATTGGTCTCTATCAGACACATAAC
 CCGTCTCAACCAATAGTATGAGCTAATGCCTCAAAAACCTCTTCTTTGCTCTGCATCTGT
 AAATCCGTCTGGATCAGACTCACATTAAGAATATCTTTGATTTCCATATATTATCTCCCG
 TAATTCTTCTTTTGTTAACTGTTTTAATTGATTTATGAATGATTCATCTGCTAGTCTTCT
 CATCAATGTTTTAATACATGACTTGTCCTGTGATACTGCAATGGCCAAACCGATAATAAG
 GTCAACACACTGGATATCCTTCGACCATTCTCTGATAGGTGGTTTTAATCTAGTAATCAC
 TAAGACATGATGTTGAAAGTTTCCTTCACAATGTGGTAGAAGAACACCTTTAGCAACCTC
 TATACTTCCCTGTCTCTCACGGTAATATAGAAGCTCTTCTATTTTTTCTGTATCTTCAGA
 AACAGAAGGCTGATTTGATTTGCTAATTCTTTGTAGGCTTCTTGACGATTTTGAACAGA
 TATATCCATAAGGACAAGCGAAAGATTATTCATAGTTTATCTCCTGAATTTTTTGCTTGAA
 GACGTTGTTTATCACCTCGGTTAGAAAAGCACTAACTAGGACAAACGGGACACTTGCTG
 GTTCCTGCAAAGCTACCGTCGTCACAATGAAATCTAAATCTGGATATAGATTTATCAG

ORF Predictions:

ORF #	Start	End	Direction	Length
1	317	550	R	78 aa

> 3864582-1 ORF translation from 317-550, direction R
VEKAGVVIAINHNEIPWETIDGKGVKVIVLFAVGDDTEAAREHLKTLSLFARKLGNDEVV
AKLVRAQTSDDVIAAFC*

Description:

Probable phosphotransferase enzyme IIa component

Assembly ID: 3864604

Assembly Length: 2077bp

> 3864604 Strep Assembly -- Assembly id#3864604

CTAGTCTTGCTACTGTCTAAGTTGGCTTGTGCATAAGCCTGCCAGATTTTTTGTGGGG
TTTGGCAAGTGGGTAATTCTTGAATTCTTCTGGTGAAAGCCAACGAACCTTCCCTATCTGA
AAAATCATGGAAGTCACTCACCTGACCTGCTACAATCTGTACATGCCATTTTCGATGACT
AAAAACATGCTGGACTGTATCAAAACAAACATCAAGCCAATCAACATCTAGGTCATAGTC
CTGCTGGAACTCTCTTCTGGGACTGGGGCCAGAGTTCACACTTTCTTCCGCAACCTGAT
GAAAGAGGTCAAACCTGCTCTTCTTGCGAAAAGTTATCAACTTCTATAAAGGGGAAATGCC
AAAAACCTGCCAAGAGCTTTTCGCTTTCATTTTTTTCAAGTAAAAATGTCTTGAGAAT
TTTTTCACTAAGGCTTTAAGATAAATAGGAACCGGCTTTTTCTTAGGAGATTTAATTG
GATAACGGTCCATGGTTCCATTCTGATATGCCGCACCTAAAGTCCTTGACTGGGCTTTCTT
CAGGTCTGGGATTTACAGGAGACTCAATATCAGACCCTAAGTCCATCAAGGCTTGATTAA
AATCACCCGGACGATCTGGATTAATCAAGATCTCCATCATTGCCTGAAAAATTTTTCGAT
TACTTGGAATCCCAATATCGTGGTTGACTTCAAACAGACGCGCCAAGACCCGCATGACAT
TACCATCTACAGCTGGCTCAGGCAAGTTAAAAGCAATACTGGAAATGGCTCCTGCTGTGT
AAGGTCCAATCCCTTTCAAGCTGGAAATTCCTTCATAGGTATTTGGAAATGGCCACCAA
AGTCAGTCATAATCTGCTGGGCTGCAGCCTGCATATTGCGAACTCGAGAATAATAACCCA
AGCCCTCCCAAGCTTTTCAGTAAACTCTCCTCAGGCGCAGTTGCCAGACTTTTCGACAGTTG
GAAACCAGTCCAAAAATCTTTTCGTAGTAAGGGATAACTGTATCCACCCTGGTCTGCTGAA
GCATGATTTTCAGATACCCAGATGTGATAAGGATTTTTTACTTCTCCTCCAAGGCAAATCTC
TTTTGTTTTTCATCATAACCAAGCGAGAAGTTTTCTCACC GGAAAGAAATGACTTTCTCCTC
CGGCCACATGACGATACCGTATTCTTTCAAATCCTAACATATCTCTAGTTATAACACAGA
AGGTTTCACCTGTCTTTGTATCTGATTTATAATATTTTCAATAGATAGTATATAACTTTT
CCTATCTACTTATACTCCAATGAAAATCCAAAGAGCAAACCTAAGAAGCTAGCCGCAGGTT
GCTCAAAACACTGTTTTTGAGGTTGTGGATAGAACTGACAGAGTCAGTATCATATTACCTA
CGGCAAGGTGAAGCTGACGTAGTTTGAAAAGATTTTTCGAAGAGTATAAATCTTATTGATG
AACTGCTTGCAGTCTGAGAAAAAATGAGCTTGGATATTATTTCCAACTCACTTAAAGTC
AATTTCAATCCACTAGAACAAAGCCTAGTACAGTTCCATCGCTTTCAACATCCATGTTGAG
AGCTGCTGGACGTTTTTGGAAGACCTGGCATGGTCATAACATCACCAGTTAAGGCAACGAT
GAAGCCTGCACCTAATTTTGGTACCAATTCACGAATGGTAATTTCAAAGTTTTCTGGGTGC
TCCAAGCGCATTTGGATTGTCTGAGAACTGTATTGAGTTTTAGCCATACAAATTGGCAA
TTTGTCCCAACCGTTTTGAACGATTTGAGCAATTTGTGTTTGAGCTTTCTTCTCAAAGTT
CACTTTGCTACCACGATAGATTTTCAGTGACAATTTTTTTCAATCTTTTCTTGACAGAAAG
GTCATTATCGTACAAACGTTTATAGTTAGCTGGATTTTCAGCAATTGTCTTAACAACCTGT

TTCGGCAAGTGCTACTCCACCTTCTGCTCCATCAGCCCAGACACTAGCCAATTCAACTGG
TACATCGATTGAGGCACAGAGTTCTTTTAAGGCTGCAATTTTCAGCTTCTGTATCAGATAC
AAATTCGTTAATAGATACAAGCTAATGGAATACCGAA

ORF Predictions:

ORF #	Start	End	Direction	Length
1	1	141	R	47 aa
2	1513	1803	R	97 aa

> 3864604-1 ORF translation from 1-141, direction R
VSDFHDFS DREVRWLSPEEFKNYPLAKPQQKIWQAYA QANLDSSQD*

Description:

unknown

> 3864604-3 ORF translation from 1513-1803, direction R
VNFEKKAQTQIAQIVQNGWDKLPICMAKTQYSFSDNPNALGAPENFEITIRELVPKLGAG
FIVALTGDMVTMPGLPKRPAALNMDVESDGTVLGLF*

Description:

FORMATE--TETRAHYDROFOLATE LIGASE (EC 6.3.4.3)
(FORMYL TETRAHYDROFOLATE SYNTHETAS E) (FHS) (FTHFS). -
CLOSTRIDIUM ACIDI-URICI.

Assembly ID: 3864610

Assembly Length: 1887bp

> 3864610 Strep Assembly -- Assembly id#3864610
CTCAAAACNCTGCTTTGAAGAGATTTTCAAAGAGTACAAGAAGTTTAGTTATTAGCGTTC
TTACCGCTTGTAAGTAGATTTCTCATAAAATAGAATCTTTTCCTTTTAGTTGTAACTA
GTCTGGGAGAGTAGAGAGGTTTGAGATACCTTTCTAGCTTTTGGAATTATCATCTAAGAAG
AGTAATTTCCCTTGCAATTAAGGGGAGAGACACGAAATGACTATAATGGGTGAC
AATGGGGGAAGGGATAGACAAGAGATTTTATCCACATATGAAAAAGGAGGTTAGGAAAG
AGTTATATATCCTATATTATATAAATAATCAATTGCGCAGAAATTTGGTAAGAATTCATG
CGTCAACTCATAAAGAACTACTTAAAAAATTCACAGTATTCATAATTATTTTCGAGGAGA
AAAACAGTGAAAAAAGAAAAAGCTTGCTCTGTCTCTTATCGCTTTTGGCTGACGGCT
TGTTTAGTAGGCTGTGCTAGCTGGATTGATCGTGGAGAATCCATAACGGCTGTTGGCTCA
ACTGCCTTGCAACCCTTGGTTGAAGTAGCGGCAGATGAATTTGGCACCATCCATGTTGGA
AAAACGGTCAATGTCCAAGGGGGAAGTTCTGGTACAGGCTTGTCACAGGTTTCAGTCTGGG
GCAGTTGATATAGGAACTCAGATGTATTTGCTGAGGAAAAAGACGGAATTGATGCTTCT
GCTCTTGTTGACCACAAGGTCGCGGTAGCTGGCTTGGCTCTGATTGTCAATAAGGAGGTT
GATGTTGATAACCTAACGACAGAGCAACTTCGTCAAATCTTCATAGGTGAGGTAACCAAT

TGGAAAGAGGTTGGTGGTAAGGACTTACCCATCTCTGTTATCAATCGGGCAGCCGGCTCT
 GGCTCTCGTGCTACCTTTGATACTGTCATTATGGAAGGTCAGTCTGCCATGCAAAGTCAG
 GAGCAGGATTCAAATGGAGCGGTAAAATCAATCGTATCAAAAAGTCCAGGAGCTATCTCT
 TATTTATCTCTTACCTATATAGATGATTTCGGTCAAAAGCATGAAGTTGAATGGCTATGAC
 TTAAGTCCAGAAAATATAAGTAGCAATAATTGGCCCTTGTGGTCTTATGAGCATATGTAT
 ACATTGGGGCAGCCCAATGAGTTGGCTGCAGAATTTCTCAATTTTGTTCCTCTCGGATGAG
 ACCCAAGAAGGGATTGTCAAAGGATTGAAGTATATTCCGATTAAGGAAATGAAGGTTGAA
 AAAGATGCTGCCGGAAGTGTGACAGTGTGGAAGGGAGACAATAATGAATCAAGAAGAAT
 TAGCTAAGAAAATGTTGCTTCCATCAAAGAATTCTCGTCTGGAGAAATTAGGAAAAGGTT
 TGACCTTTGCCTGTCTTTCTTTGATAGTCATCCTTGTGGCCATGATTTTGGTTTTCGTAG
 CGCAAAAAGGCTTGTGACCTTCTTTGTCAATGGTGTGAATATCTTTGACTTTCTTTTGG
 GAGGAACTTGAATCCTTCTAGTAAAGAATTTGGTGCCCTTCCTATGATTTTGGGTTCCCT
 TTATCGTTACCATTCTCTCAGCCCTTATCGCAACACCCTTTGCTATTGGTGCAGCAGTTT
 TTATGACCGAAGTATCACCAAAAGGGCGAAGATTTTGCAACCAGCTATTGAACTCCTGG
 TTGGGATTCCCTCAGTAGTGTACGGATTTATTGGCTTGCAAGTCGTCGTTCCCTTTGTTC
 GCAGTGTCTTTGGTGGGACTGGTTTTGGGATTTTGTGAGGATTTCCGTCCTCTTTGTCA
 TGATTTTGCCGACCGTAACCTTTATGACAACGGATAGCTTGCGTGCGGTTCCCTCCNTTAT
 TATCGTGAAGCCAGTTTCGCTATGGGA

ORF Predictions:

ORF #	Start	End	Direction	Length
1	427	1305	F	293 aa

> 3864610-1 ORF translation from 427-1305, direction F
 VKKRKKLALSLIAFWLTACLVGCAWIDRGESITAVGSTALQPLVEVAADDFGTHVGKT
 VNVQGGSSGTGLSQVQSGAVDIGNSDVFAEEKDGIDASALVDHKVAVAGLALIVNKEVDV
 DNLTTTEQLRQIFIGEVTNWKEVGGKDLPISVINRAAGSGSRATFDTVIMEGQSAMQSSEQ
 DSNGAVKSIVSKSPGAISYLSLTYIDDSVKSMKLNGLYDLSPENISSNNWPLWSYEHMYTL
 GQPNELAAEFLNFVLSDETQEGIVKGLKYIPIKEMKVEKDAAGTVTVLEGRQ*

Description:

PROBABLE ABC TRANSPORTER BINDING PROTEIN PRECURSOR (ORF108). -
 BACILLUS SUBTILIS. (BLAST)

Assembly ID: 3864716

Assembly Length: 405bp

> 3864716 Strep Assembly -- Assembly id#3864716
 CTGAGGAATCAAAAGTTGAACCACCAGTAGAACAAGCATAAGTCCCAGAACAACCCGTGC
 AACCTACACAAGCTGAGCAACCAAGTACACCAAAAGAATCATCACACAAGAAAATCCTA
 AAGAAGATAGGGGAGCGGAAGAGACTCCGAAACAAGAAGATGAACAGCCAGCAGAAGCCC
 AAGAAATCAAGGTTGAAGAACCAGTAGAATCTATAGAGGAGACTGTCATTCAACCTGTTG

AACAACCAAAAGTGGAAACGCCTGCTGTTTAATAACTAACGGAACCTACAGAGGAACCTA
 AAGTTGAAGTAACTAGTATTCCCCTCACTACTCGCTATGAGGAAGACCTTACTTACGAAC
 ACGGAACGCGTTGAAGTTGTTAAGGAAGGTTATAATTGGCAGTAT

ORF Predictions:

ORF #	Start	End	Direction	Length
1	57	272	F	72 aa

> 3864716-1 ORF translation from 57-272, direction F
 VQPTQAEQPSTPKESSQQENPKEDRGAEETPKQEDEQPAEAQEIKVEEPVESIEETVIQP
 VEQPKVETPAV*

Description:

unknown

Assembly ID: 3864718

Assembly Length: 1542bp

> 3864718 Strep Assembly -- Assembly id#3864718
 CTATGGGATTGGTAGTTCCTTAGTGCAGGGGCTGTAGACCCAGTTGCGACCCTAGCGC
 TGGACTAGTCGAGAGGGTGTTGTTGAAAATGGATGGCTATCGCTATGTTGGTTATCTATC
 AGGTGACATCCTCAAAACGCTTGGCTTGGACACTGTTTTAGAAGAAACCTCAGCAAAACC
 TGGAGAGGTGACTGTAGTCGAAGTTGAGACTCCTCAATCAACAACAAATCAGGAGCAAGC
 TAGGACAGAAAACCAAGTAGTAGAGACAGAGGAAGCTCCAAAAGAAGAAGCACCTAAAAC
 AGAAGAAAGTCCAAAGGAAGAACCAAAATCGGAGGTAAAACCTACTGACGACACCCTTCC
 TAAAGTAGAAGAGGGGAAAGAAGATTCAGCAGAACCATCTCCAGTTGAAGAAGTAGGTGG
 AGAAGTTGAGTCAAAACCAGAGGAAAAAGTAGCAGTTAAGCCAGAAAGTCAACCATCAGA
 CAAACCAGCTGAGGAATCAAAAGTTGAACCACCAGTAGAACAAGCAAAAGTCCCAGAACA
 ACCCGTGCAACCTACACAAGCTGAGCAACCAAGTACACCAAAAGAATCATCACAACAAGA
 AAATCCTAAAGAAGATAGGGGAGCGGAAGAGACACCGAAACAAGAAGATGAACAGCCAGC
 AGAAGCCCAAGAAATCAAGGTTGAAGAACCAGTAGAATCAAAAGAGGAGACTGTTAATCA
 ACCTGTTGAACAACCAAAAGTGGAAACGCCTGCTGTAGAAAAACAAACGGAACCAACAGA
 GGAACCAAAAGTTGAAGTAACAAGTATTCCCCAAACTACTCGCTATGAGGAAGACCTTAC
 TAAGGAACACGGAACGCGTGAAGTTGTTAAGGAAGGTAAGAATGGCAGTAGAACAGTTAC
 TACTCCATATATCTTGAATGCGACAGATGGTACGACTACAGAAGGCACTTCGACAACCTGA
 TGAAGCTGAGATGGAGAAAGAGGTTGTTTCGTGTTGGCACGAAACCCAAAGAAAAATTAGC
 TCCAGTCTTAAGTTTGACAAGTGTTACAGATAATGCAATGTTGCGTAGTGCGAGACTTAC
 TTATCATTTGGAAAATACAGATAGTGTTGATGTGAAAAAAATTCATGCTGAAATTAAAAA
 TGGCGATAAGGTTGTCAAACTATTGACTTATCTAAAGAGAGATTATCAGATGCTGTTGA
 CGGTCTTGAACCTTATAAAGATTATAAGATTGTGACGAGTATGACCTATGATAGAGGTAA
 TGGTGAAGAAACCTCTACGTTGGAAGAACTCCACTACGATTAGACCTCAAGAAGGTTGA
 ATTGAAAAACATCGGCTCTACTAATCTCGTCAAAGTAAATGAGGATGGTACTGAGGTGGC

AAGTGA~~CTTCTTAACAAGTAAACCTGTGGATGTGCAGAATTACTACCTCAAAGTAACTTC~~
 CCGTGATAATAAAGTTGTTTCCCCTCCCAGTTGAAAAAATTGAAGAGGTGACTGAGGAAG
 GTCCACCACTTTACAAAGTCCCTGCTAAGGCCCTAATTTGAT

ORF Predictions:

ORF #	Start	End	Direction	Length
1	77	1474	F	466 aa

> 3864718-1 ORF translation from 77-1474, direction F
 VLLKMDGYRYVGYLSGDILKTLGLD~~TVLEETS~~AKPGEVTVVEVETPQSTTNQEQARTENQ
 VVETEEAPKEEAPKTEESPKKEPKSEVKPTDDTLPKVEEGKEDSAEPSVVEEVGGEVESK
 PEEKVAVKPESQPSDKPAEESKVEPPVEQAKVPEQPVQPTQAEQ~~PSTPK~~ESSQQENPKED
 RGAEETPKQE~~DEQPAEAQEIKVEEP~~VESKEETVNQPV~~EQPKVETPAVEKQTEPTEEP~~PKVE
 VTSIPQ~~TTRYEEDLTKEHGTREVVKEGKNGS~~RTVTTPYILNATDGT~~TTEGTSTTDEAEME~~
 KEVVRVGTKPKEKLAPVLSLTSVTDNAMLRSARLTYHLENTDSVDVKKIHA~~EIKNGDKV~~V
 KTIDLSKERLSDAVDGLELYKDYKIVTSMTYDRNGEETSTLEETPLRLDLKKVELKNIG
 STNLVKVNEDGTEVASDFLT~~SKPVDVQNYLKVTS~~RD~~KNKV~~SPPS*

Description:

unknown

Assembly ID: 3864802

Assembly Length: 1321bp

> 3864802 Strep Assembly -- Assembly id#3864802
 ATCGAATTACTTCAACTCCA~~ACTTTACTCTCAATAAAAATCAAATGTAAAAAGAGGAGCT~~
 AAATTTATCTTTTTTCTCCTCCTTCATCGTTCTTACTTTTGACCATAATAAGCATTTGGTC
 CATGTTTACGTTGGTAGTGTTTTTCTAGTATGTACTGGGGAGCAGGTTCAACTCTTGGAT
 TGATTTGTTCTGTAAAGCGATT~~CATCTTTGATACTTCCTCTAGTACGACAGAGTGATAAA~~
 CAGCATTTCTCTGGATTTTTGCCCCAGGTGAATGGACCGTGATTGCGTACAACAATTCCTG
 GTACTTCAACCGGGTTAAGTCCGCGATGTTCAA~~ACTCTTCTACGATAACCAGGCCAGTAT~~
 CTTTTTCATAGGCCACTTCTACTTCGTCTTGGTCAA~~ACTACGGGCGCAAGGGATTGAAC~~
 CGTAGAAATAATCTGCATGGGTTGTTCCGTAGAAAGGAATATCACGACCTGCCTGAGCCC
 AAGCAACAGCTTCTGTCTGAATGGGTGTGAACCACACTACCAATTTCTGACCAAGCCTTAT
 ATAATTGCACATGAGTTGGGAAGTCGGAAGATGGTCTTAAATCCCCTTATAGGATCTTAC
 CATCTAGATCAGTCACTACCATGTTTTTCAGGTGTCAATTCGTCATAATCCACGCCTGATG
 GTTTGATAACAATGACACCGAGTTCGCGATTGACTTCAGATA~~CATTCCCCCAGGTAAATT~~
 TGACAAGTCCATGTTTTGGCAATGATTGATTGGCATCACAGACTCGTTTTACGCATAGCAT
 TGATTACTTGATTCATCTTACATCAAACCTGCTTTCTTAATGAGTGGATAGAGAAAAGCT
 TGCGCCTCTTGAATGGCTGCGCGTGTTTCTTCTACTGTTTCACAATTTTCAGACCACATT
 TCGATTAGGAAAGGTCCATTATAATTGGTTTTCTTTAA~~AATATCGAAAGCTTCTTCCCAT~~
 TTGACACAACCTTGCCCCAAAGGTACATCTCGGA~~ACTGGCCCTTTGAACTTTCTGTCACT~~

GCATAAGTATCCTTGAGATGGAGAGTTGCGATGGCATGATGACCAAGATAAACTCACTA
 TAGATATCATTATGCCATGCAGACACATTACCAATATCTGGATATACAAAGAGGAAGGGA
 GAGTCAATCTCTTTTTCTATAGCCAAATATTTTTTCGATGCTATTGATGAAAGGATCATCC
 ATAATTTCAATAGCAAGTACCACCTGAGCTTCTTCAGCCCAGTCACAGGCTTTTCTCAA
 TTTTGTATAAAACGTTGGCGTGTCTGGGGTGACTTTTCCTCATAGTAAACATCGTAACCA
 G

ORF Predictions:

ORF #	Start	End	Direction	Length
1	92	550	R	153 aa

> 3864802-1 ORF translation from 92-550, direction R
 VQLYKAWSEIGSVVHSTHSTEAVAWAQAGRDIPFYGTTHADYFYGSIPCARSLTKDEVEVA
 YEKDTGLVIVEEFHRGLNPVEVPGIVVRNHGPFTWGKNPENAVYHSVVLEEVSKMNRFT
 EQINPRVEPAPQYILEKHYQRKHGPNAYYGQK*

Description:

L-RIBULOSE-5-PHOSPHATE 4-EPIMERASE (EC 5.1.3.4). - ESCHERICHIA
 COLI.

Assembly ID: 3864854

Assembly Length: 1265bp

> 3864854 Strep Assembly -- Assembly id#3864854
 TTTTTCTGTTTTTCGGAGCAAACCTGGGCTCCAGCCGGTTTTGGCCTTCTTTCCCTTAGCTA
 CAGCTGGTTTTAGCTGGCTCAGATTTTTTCGGCTTTCTTTTCTGCACTTACTTTTGGTGCTG
 CAGGTTTTGCTTCTACTTTTCGGAGCAGCTGCAGGCTTAAAGCTGGCAGCAATTTTTGCAG
 CGACAGCTTCTTCCACACTTGATGAGTGGCTTTTCACATCCAAGCCCAACTCTTTTGCAC
 GCGCTACAACCTTCTTTACTTTCTTTTCCAAGTTCTTTTGCGATTTCGTACAATCTTTTCT
 TAGACAAATCATGTCTCTCTTCTATTCCATAAGAGACCTCATTTTCTTTGTAAATCCA
 GCATCTGTTACAGCCAAAACCTTTCTCGATTTCCCGACTGCTATGATTAATTCCAGTGTT
 GAAAACACGGTTACAATTTCTACTTGATAATAATGACTTTTATCTTGAATCTTCTTGGTC
 AGATTGGGTCCAGCATCATGAGCTAGAAAGACCAACTTGGCCTTGCCGTCTTGAATGGCC
 TTGACCACCAATTCTTCACCCGATATGATGCGCCCTGCTCGCTGAGCAAGCCCCAAGAGA
 TTACTTATCTTTTGCTTATTCAAGTCCCAACTCTCTTCTTTTCACTTTGTGATCCACATA
 AGCGATCAACTCGTCATAAAAGCTTTCTTCCACTTCCATGCTAAAGCTGCGGTTAAAGAC
 CTTCTTCTTTTTTCGCCCTCTAGGGCTTCTGCATTGTCTAGTTTGATATAAGCGCCGCGGCC
 ATTGGCCTTGCCCGTAGGATCAATAAAGACTTGTCCTTCTTGTCTTGACAATGCGGAG
 CAAATCACGCTTATCAATCACTTCGTTAGACACAACAGACTTGCGCAAAGGGATTTTTCT
 TGTTTTCATCTTTCCCTCCTCTAGCAGCTTTTATTCTTCTACAGTATCGTTTTCTACTTC
 CAACTCTACTGAAGCAGCGTCTTCCATGGCTTCAAATTCGCTAGCAGACTTGATATCGAT
 ACGGTAACCAGTCAAGTGAGCCGCCAAGCGCACGTTTTGTCCACGACGACCAATGGCAAG

AGAAAGCTTGTTATCTGGAACAACCACCAAGGCACGTTTGCTGTCGTTTTTCATCAAAGAT
 AACTTGGTCAACCTCAGCAGGAGCGATGGCATTGTAGATAAATTCAGCTGGATCTGCTAC
 CCACTCGATAACATCGATATTTTCTTCGATTGGTACCATGCGGTCATTTTTAGCATCGTA
 ACGAG

ORF Predictions:

ORF #	Start	End	Direction	Length
1	324	548	R	75 aa

> 3864854-1 ORF translation from 324-548, direction R
 VVKAIQDGKAKLVFLAHDAGPNLTKKIQDKSHYYQVEIVTVFSTLELIIAVGKSRKVLAV
 TDAGFTKKMRSLME*

Description:

PROBABLE 11.1 KD RIBOSOMAL PROTEIN IN NUSA-INF8 INTERGENIC
 REGION (ORF4). - BACILLUS SUBTILIS.

Assembly ID: 3864862

Assembly Length: 1305bp

> 3864862 Strep Assembly -- Assembly id#3864862

ATAAACCAAAGGAAGCTGAGCTCTTTAGTCCCAGCTTCTTTTTATATATAAAATTTTACC
 CGTGAAAAGACAGGGCCTTAGCAGACTTCTTTTTTACTTCGTTCCACCCTTGCTTTTTTCTT
 TGTATGTTTGGGCGTTGGCAGTTGGTTATACATAGCTAAAATCAGGTCTTATAGAAACAT
 CTTATTATCAAGTTCTTCCACTCAAATCATTTCTTTGGCACCTTTGTATGGAAACTCAAA
 AGAAGATTGGTCAATCTTATCTAAGACTGCTTGCACGGGTTTAACTAAAAGCGATCGTCA
 TAAATGCCGCCAATAATCTTGCCGCGGAAGTAAAGAATATACTCCCCCATCATGGAACGG
 TAAGTCACATCATCTAATCCTGATAATTGTTCCAAAACAAATTCCAAATAGTTCTTACTT
 GATGCCATTTCTAATCTTCTAGGCTCTGTTCAACGATAACAACCGTATAGAGTTCTTGCT
 TAACCTCGCATCCAATTGATTTAAAGCCCTGCTTTTCCCAAAAATGCTGAGATTGCGGAT
 TTCCCTTAACATAAGCCAAACGTGCCTTTCGAAAGTTCTTAGCAAAATAAGCTAGTGCTT
 CTGTCACAATATGACTACCAATCCCTTTCCTCTGATAGGCTTGATCAACCATAAACAAC
 CAATAAAAACAGTCTCCTCATCAGGATATGCATAGACAAAATCCATAACAGCCACAAGGT
 CAAATCCATTCCAAAATCCAACAAAAAATTTATCAGCCTTAGCTTTACCTTCAGGTAGAC
 AAAGCATGTCTCTTTTACAGTTGCAAAATTTGGCTCTGGTGGACAATGCTGAAAATACA
 GAGGATTACTTTTCATATAAAGATAAAATACTTGGAATATCCTTTTTCAGTTAGTATCCTAC
 AACTGTAATACTTAGATAGTTGGTCAATCATCTTTTCAAATTCGATACTTTCTTGCGCC
 TGTGATTATGACACAGGAAGATGCACTGATCGTCATCAGCCACATAAAAGTTCTTTCCAT
 CGTGCCTAATCGTTGTCTCAAACCTTTGGATAAAACCTTTAGCCTATACAACCTGGATTTT
 CCTCTCTCAAAAGTATATTCTTTTGCAGGCGAACTTCCTCAAAATCAGTCGTGTGCAACT
 TCAGTAGAATATTCATAGGCTCGGATAATCTGAGCGACAACAGGATGGCGAACCACATCC
 TTGGCTGAAAAATGAACAAAGTCAATCTGATGGATGTTCTTGAGTTTCTCTTGAGCATCA

ATCAAACCGGACTTGACATTACGTGGCAGGTCAATCTGACTAATA

ORF Predictions:

ORF #	Start	End	Direction	Length
1	431	1003	R	191 aa

> 3864862-1 ORF translation from 431-1003, direction R
 VADDDQCIFLCHNHRAQESIEFEK MIDQLSKYYSCRILTEKDIPSILSLYESNPLYFQHC
 PPEPNFATVKEDMLCLPEGKAKADKFFVGFVNGFDLVAVMDVFYAYPDEETVFIGLFMVD
 QAYQRKGIGSHIVTEALAYFAKNFRKARLAYVKGNPQSQHFWEKQGFKSIGCEVKQELYT
 VVIVEQSLED*

Description:

unknown

Assembly ID: 3864888

Assembly Length: 1742bp

> 3864888 Strep Assembly -- Assembly id#3864888
 CTAATCTCCTTAAAACGTGATCTTTTCAAGAATATTTTATCTAAACAATCCAGCAAGTC
 TTGGTAAGAATAGACTTCGTAAGTCGGCTGGGCTTGTGTGTGATTTTCGAGGTGATGAGG
 ATTATACCAGATAGTGTCAATCCCCGCATTATTGCCACCTTGAATGTCGGCGGTAGAGA
 ATCTCCAATCATCAGCGTCTTTTCTTTACTAAATCCAGCAATTTGCTGGCCAATCTTTTC
 ATAAAAAAGAGCATCCGGCTTTTGAGTTTGCAACTGTTCTGAGATAAAGACTTGATTGAA
 ATAAGGTGCTAGACCAGATTGAGCCAAACGTCCTGTCTGAATGGCAGTAATGCCATTTGT
 CGCAGCATACAAGTTATAATCACGCTCAATGAGGCTGTCCAAGAGATCATGAGCGCCCGA
 TAGTGTTTGTCCCTGCTGGGCGAGGTAAAATTGGTAACGCTGGGCAAGAAAACCTACCGTC
 TTTTTCCTGTCCAAAATGAGCAAATAAACGAGAAAAGCGCGTGTTAACCAGCTCTTGTTT
 ACTGATTTTCTTCAGCTCCAAGTCTTTCCAGAGAGCCTTGTTTCATAGGAACGTAATAATC
 TTTATAAGCCGGAATATCCGCAACTCCTTCTTCTTTTAGAAGTGGAGTCAAAGCCACATC
 CTCAGCAGCATCAAAATCAAGAAGAGTGTGGTCGAGGTCGAAGAGTACAAATTTGTAGAA
 CAATTTGAGGTTTTCTTTCTGAAAATTCATTAAGAACATTATATCATAAAGCACCTCAT
 ACAATTAATAATTTAATCACTTAAAAAAAATTCGAACACTTTCTATACAACCTGACAGCT
 CAAATCTTTCAGAATAGAACAATACTAACTATCGAACACCCCGTCTTCATAAATACATAT
 GTAATTCTAGGCCTAGAATTCCCTATAAACTAAATGCTTTCATACTCTTCCAAGTAATTGA
 TTGCCTTAAATTTTAATTTTGAAGGTTTCTAAAGCTAGAATAGCCCCATCACAATCAGT
 TTTGATTGATTCACAATTTAGAAACACTATAGTTTCACTCCTGTAAATAAAAAGGAAC
 TGCATAAAGCAATCCCTTTCTGATTTTGAATCATTTACTTAACATTTTATAGTTGAGAT
 AATCAATAGCTTATCTATAAAAAGAGTTATAGTAAATTCCTTATTTATTGATTCCAAGC
 TCCGCTAACTGTATTTGAATAACTGACAGTTCTGCACCAGCCTGAAAAAGAGCAGCTGCA
 TTATAGGCACCTTCTACAATTGGAACCCTGTTGATGATGATACTTTTATCACTGAAATCA
 GTCACCATTTTAAAGTTCATTTTAGCAGAACCTAGGTCAAAAAAGGCAAGTAAAGTATCT

GCTGGATTTTCGGAAACAACCCTATCTACTTGATCAAAACTCGTTCCAATTCCTCCGCC
 TCGGTTCTCTACATAAGTAATCGGAACATCTTTAGCTACTTTACTAATCAGTTCAACA
 ACACCTTCTGCAATGTGTTTGGAAATGTGAAACGATAACAAGACCAATACCAATACTTTCC
 ATCAAACCACTCCAGTTTCTAAAATAGCAGTAAAGAGTAATCCTGATGAGAATGATCCAG
 GATCAATATGTCCAAGAAACCACATGCTCCTAAGACAAGAGCTAACAGACTGGCCATCAA
 TAATAGTATTGTTCTTTTTTTCATCATTACTCCTTAAGTAGTGTTTAACTGATTAATTCG
 AT

ORF Predictions:

ORF #	Start	End	Direction	Length
1	10	657	R	216 aa

> 3864888-1 ORF translation from 10-657, direction R

VALTPLLKEEGVADIPAYKDYVPMNKALWKDLELKKISKQELVNTRFSRLFAHFGQEKD
 GSFLAQRYQFYLAQQGQTLSGAHDLLDSLIERDYNLYAATNGITAIQTGRLAQSGLAPYF
 NQVFISEQLQTQKPDALFYEKIGQQIAGFSKEKTLMIGDSLADIQGGNNAGIDTIWYNP
 HHLENHTQAQPTYEVYSYQDLLDCLDKNILEKITF*

Description:

unknown

Assembly ID: 3864898

Assembly Length: 1136bp

> 3864898 Strep Assembly -- Assembly id#3864898

GTGGAATGCGGGGACGCCTTGTCTAATTTTGGATCAAGCCCTGAGTTTGACACAGGGAAA
 TGAGCTGGACGGACTGCTATCTCTGAAGAAATTACTGGCACCATTAGCCTATCAGCCTTG
 GATGATTATGTGGCGGCCTTGTCTCAACAGGATGTTCCCAAAGCTTTGTCTTGCTTGAAT
 CTTCTTTTGTGACAATGGTAAGAGCATGACTCGTTTTGTGACCGATCTTTTGCACATTTA
 AGAGACTTGTTAATTGTTCAAACAGGGGGAGAAAATACTCATCATAGTTCAGTCTTTGTA
 GAAAATTTGGCACTTCCTCAAAAAAATCTGTTGAAATGATTCGCTTAGCAACAGTGAAT
 TTAGCAGATATTAAGTCTAGTTTGCAGCCCAAGATTTATGCTGAAATGATGACCGTCCGT
 TTGGCGGAAATCAAGCCCGAACCAGCTCTATCAGGAGCGGTGAAAATCGAATTGCTACG
 CTGAGACAGGAAGTTGCCCCGTCTCAAACAAGAGCTTTCTAATGCAGGTGCGGTTCCCTAAA
 CAAGTTGCACCAGCTCCTAGTCGACCAGCTACGGGCAAACAGTCTATCGTGTGCGATCGC
 AATAAAGTGCAATCTATCTTACAAGAGGCCGTGAAAATCCTGATTTAGCACGTCAAAAT
 CTAATTCGTTTGCAGAATGCCTGGGGAGAGGTAATTGAAAGTCTAGGTGGGCCGGACAAG
 GCTCTGCTAGTTGGTTCTCAACCGGTTGCTGCCAATGAACACCATGCTATTCTTGCTTTT
 GAGTCTAACTTCAATGCTGGTCAAACATATGAAACGAGACAATCTCAATACCATGTTTGGT
 AATATCCTCAGTCAGGCGGCAGGTTTTTTCACCTGAGATTTTAGCTATTTCCATGGAGGAA
 TGGAAGAAGTTGCGCGAGCCTTTTCAGCCAAAGCCAAATCTTCTCAAACCTGAAAAGAA
 GTAGAAGAAAGCCTGATTCCAGAAGGATTTGAATTTTGGCTGATAAAGTGAAGGTAGAG

GAAGACTAAAGAAAGATTTTCATGATACAATAAGTTTATGAATAAAACAACAATTTATTATT
ATGGCGCTATTTACAGCTGCTGAGACCTATTTTTTCAATGAAGCCTGGATGACTGG

ORF Predictions:

ORF #	Start	End	Direction	Length
1	130	1029	F	300 aa

> 3864898-1 ORF translation from 130-1029, direction F
VAALSQQDVPKALSCLNLLFDNGKSMTRFVTDLLHYLRDLLIVQTGGENTHHSSVFVENL
ALPQKNLFEMIRLATVNLADIKSSLQPKIYAEMMTVRLAEIKPEPALSGAVENRIATLRQ
EVARLKQELSNAGAVPKQVAPAPSRPATGKTVYRVDRNKVQSILQEAVENPDLARQNLIR
LQNAWGEVIESLGGPDKALLVGSQPVAANEHHAILAFESNFNAGQTMKRDNLNTMFGNIL
SQAAGFSPEILAISMEEWKEVRAAFSAKAKSSQTEKEVEESLIPEGFEFLADKVKVEED*

Description:
unknown

Assembly ID: 3864938
Assembly Length: 1670bp

> 3864938 Strep Assembly -- Assembly id#3864938
CTGTCTCTGAAACAGTCACATCAAGTGCCTCTGAACAANCGCCCNCTAGGTNGACGGT
ATCGATAAGCTCGATCTGTGATTTTCAGAGAAGAAATCAAGTGCTGTAAACAGAAGTAAGAT
GTAATTGTATGTAAAGGAGACGTCATGTTAAATAGTATTGTAACCATTATTTGTATTGCC
CTTATCGCGTTTATCTTGTTTTGGTTTTTCAAAAAGCCTGAAAAATCTGGACAAAAAGCC
CAGCAAAAAAACGGATACCAAGAGATTTCGAGTGGAAGTCATGGGAGGCTATACTCCTGAG
TTGATTGTCTCAAGAAATCAGTGCCAGCCCGCATTGTCTTTGACCGCAAGGATCCTTCA
CCATGTCTGGATCAAATTGTTTTTCCAGATTTTGGTGTACATGCGAACCTGCCAATGGGG
GAAGAGTATGTAGTGGAATCACGCCTGAACAGGCTGGAGAGTTTGGCTTTGCTTGTGGT
ATGAACATGATGCACGGCAAGATGATTGTAGAGTAGGTGGAGACTATGACAGAAATTGTG
AAAGCAAGCTTAGAAAATGGCATTTCAAAAAATCCGTATCCGAGCTGAAAAAGGCTATCAT
CCAGCCCATATCCAGCTTCAAAGGGAATTCCAGCTGAGATTACCTTTCATTCTGTCTAC
TCCTTCAAACGTGTTATAAGGGAAATTCTGTTTGAAGAAGAAGGTATCTTGGAAGCAATCG
GCGTAGATGAGGAGAAAGTCATTCGTTTTACACCTCAAGAATTAGGGAGACATGAATTTT
CTTGTGGCATGAAGATGCAAAAGGGAAGCTATATAGTCGTTGAGAAGACTCGAAAATCTC
TATCTCTCCTGCAAACGTTTTTGGATTACTAGTATCTTTACTGTGCCTCTTGTGATTCTC
ATGATTGGGATGTTGGCAGGTAGCATTAGTCATCAAGTCATGCATTGGGGAACCTTTTTA
GCAACAACGCCTATTATGTTAGTTGCGGGTAAGCCATATATCCAGAGTGCTTGGGCCAGT
TTTAAAAAGCACAAATGCCAACATGGATACCTTGGTTGCGCTGGGAACCTCTAGTGGCTTAT
TTCTATAGCCTAGTTGCTCTCTTTGCTGGTCTCCCTGTTTACTTCGAAAGTGCTGGATTT
ATCCTCTTTTTTCGTTCTTTTGGGAGCAGTTTTTGGAGGAAAAAATGAGGAAAAATACGTCC
CAAGCTGTGGAGAAATTACTGGACTTGCAAGCTAAAACCGCAGAAGTCTTGAGTGATGAT

AGTTATGTCCAAGTTCCTTTGGAACAAGTCAAGGTACGCGACCTTGATTCCAGTGCGTCC
 CGGTGAAAAGATTGCTGTTGATGGTGTCTAGTAGAAGGTGTCTCTAGTATTGACGAATC
 CATGGTGACAGGTGAGAGTCTGCCTGTGGACAAGACAGTTGGAGATACTGTCATTGGCTC
 AACCATCAATCATAGTGGAACGCTTGTCTTTAGAGCAGAAAAAGTTGGCTCAGAGACTGT
 TTTGGCTCAGATTGTAGATTTTGTGAAGAAAGCTCAGACAAGTCGTGCGCCGATTCAGGA
 CTTGACGGATAAGATTTTCAGGGATTTTGTCCAGTAGTTGTCATTTTAGGAATCATGAC
 CTTTTGGGTTTGGTTTCGTCTTGCTCAGGGATAGTGTGGTTCGTGCTTGGAG

ORF Predictions:

ORF #	Start	End	Direction	Length
1	883	1326	F	148 aa

> 3864938-2 ORF translation from 883-1326, direction F
 VPLVILMIGMLAGSISHQVMHWGTFLATTPIMLVAGKPYIQSAWASFKKHNANMDTLVAL
 GTLVAYFYSLVALFAGLPVYFESAGFILFFVLLGAVFEEKMRKNTSQAVEKLLDLQAKTA
 EVLSDDSYVQVPLEQVKVRDLDSASR*

Description:

ATCS_SYNP7

Assembly ID: 3864956

Assembly Length: 1252bp

> 3864956 Strep Assembly -- Assembly id#3864956

ACAAGAACAATTGGAACAGGTACAGGCTGTTAAAAAATCGATTAACACAGCTAGTGAAGA
 AGTGAAAAACCAAGTCTTGCTACCCATGGCTGATCACTTAGTGGCTGCTACTGAGGAAAT
 TTTAGCGGCTAATGCCCTCGATATGGCAGCGGCTAAGGGGAAAATCTCAGATGTGATGTT
 GGATCGTCTTTATTTGGATGCAGATCGTATAGAAGCGATGGCAAGAGGAATTCGTGAAGT
 GGTGCTTACCAGATCCAATCGGTGAAGTTTGTAGAAACAAGTCAGCTTGAAAATGGTTT
 GGTTATCACAAAAAACGTGTAGCTATGGGGGTCATCGGTATTATCTATGAAAGCCGTCC
 AAATGTGACGTCTGATGCGGCTGCTTTGACTCTTAAGAGTGGAATGCGGTTGTTCTTCG
 TAGTGGTAAGGATGCCTATCAAACAACCCATGCCATTGTACAGCCTTGAAGAAGGGCTT
 GGAGACGACTACTATTATCCAAATGTGATTCAACTGGTGGAGGATACTAGCCGTGAAAG
 TAGTTATGCTATGATGAAGGCCAAGGGCTATCTAGACCTTCTCATTCCCTCGTGGAGGAGC
 TGGCTTGATTAATGCAGTAGTTGAGAATGCCATTGTGCCTGTTATCGAGACAGGAACTGG
 GATTGTCCATGTTTATGTGCGATAAGGACGCAGATGACGACAAGGCACTGTCTATCATCAA
 CAATGCCAAAACCAAGTCGTCTTCTGTCTGCAATGCCATGGAGGTTCTGCTGGTTCATGA
 AGACAAGGCAGCAAGCTTCCTTCCTCGCTTGGAGCAAGTGCTGGTTGCAGATCGAAAAGA
 AGCTGGGTTGGAACCAATTCAATTCCGCCTAGATAGCAAAGCAAGCCAGTTTGTTCAGG
 TCAAGCTGCTCAAGCACAAGACTTTGATACCGAGTTTTTAGACTATATTCTAGCTGTTAA
 GGTGTTGTGAGCAGTTTAGAAGAAGCGGTTGCGCATATTGAATCCACAGTACCCATCATTCG
 GATGCTATTGTGACGGAAAATGCTGAAGCTGCAGCATACTTTACAGATCAAGTGGACTCT

GCAGCGGTGTATGTTAATGCCTCAACTCGTTTCACAGATGGAGGACAATTTGGTCTTGGT
TGTGAAATGGGGATTTCTACTCAGAAATTGCACGCGCGTGGTCCAATGGGCTTGAAAGAG
TTGACCAGCTACAAGTATGTGGTTGCTGGTGATGGGCAGATAAGGGAGTAAG

ORF Predictions:

ORF #	Start	End	Direction	Length
1	1030	1251	F	74 aa

> 3864956-2 ORF translation from 1030-1251, direction F
VTENAEAAAYFTDQVDSAAVYVNASRFTDGGQFGLGCEMGISTQKLHARGPMGLKELTS
YKYVVAGDGQIRE*

Description:

gamma-glutamyl phosphate reductase (proA) homolog - Haemophilus
influenzae (str ain Rd KW20)

Assembly ID: 3864958

Assembly Length: 1785bp

> 3864958 Strep Assembly -- Assembly id#3864958

CTGCCCTAGCAGGAACGCAAGAAGGAAGTGGAGAATAGGCATTTTCAAATTTATAACCTA
CACTAGCCATCATATCTAATGTTGGAGTGCTAACTAGCTTATCCTTACTATTCAAGGATA
AGGCGTCTGCTCTCATTTGATCTACAACAATCAAATAATATTTGGTTGTTTTGTCTGAA
CCATAAAATCTCCTTTCTAATATGGCAAAAGAGGCACAAGAAGATATCTACCTTTACTGC
ACCCCTTTCTATATCAATCTCTCTATATAAAGCAATAACATTCTTGTTATGTTTTATAGA
ACAATGGACTAAAATATGACTAAATCGATTAGGAAATTCAAATCATTTTCTAGTACTGTT
TTAGTAAGTTACAGTGTACTATTCCAACCTTCAATAAATTATAAACCTTTGTCTAATAACA
ATTTTAGTGGAGATAAGAAATCCTACACCTAACTCATCTTACACGTAATCTATTTCTATT
TTATCACAAAAAACGCAAGTAAGACCATTAACCTCAATTCAGTTTTATCTGCCATTTTCAC
AAATGGGAAATAAGTCAAGACACTAATAATCAAACAAACAAGTATAAGATGATGGCACG
CCAATCAAATGCTGTAGAGAAGAAACCATATAAAATTTGGAGGCATTACCCAAGTAACATT
TTGTGTAACAGGTGAAACAAGACCCAGCTTGTTGCCAGTAAGCTACCGTTGCCATGAA
AACCAGGCTAAGTACAAATGGTATAAATAGCAAAGGATTCAAGACAAGTGGTAAACCATA
ATTGATACCGGCTCACCAATATTAAACAGAACTGGTGCTAGACCAAGTTTAGCAACTTT
TCGATAATGACTGTTTCTTGAAAAAATTAAATAGCAAGTACTAATCCTAATCCTCCAAA
CCAGACAAACGCCCCAAAAGACCCACTTGTCATATATAAGGAATCGGTTTCACCTTTTTG
GAAAGCATCCAGATTGCTAACATAGCAACTCCAAATAGCCCTTCCATGATGGGAGCCAA
TACATTTCTCTCCATGGAGACCAAAAAACCAGAATAACTTATTCAAAAAGATCATCAGAAT
AACTGCAAAGAACTTTGAGACAAACCTAGTAATGGCGTTTGTAACACCTTGTAACCCCA
ATCAATCAATAAGTCATTGCTAAGTAAATGGAAAACATAAGTCAAGATGGCTACTATATA
CATCGCCATAAATCCTGGAATGATAGAAGTGAACGGCTTAGCAATCGCAGGGGGAACTGA
ATCTGGTAACTTGATTACCCAGTTCTTTTTTCACTTTTACAGAAAATAATAGAGGCTAA

AAATCCAATCATCATGGCTGTAAAGTAGCCTCTGGCATTAATATGGTTTCCTGGAATCAC
 ATTCCCAATAGTTACCATCAGATTTTTTACCATCAAATGCTAGATTATCAATTCCATGTTA
 AGATTTGATCTAATTTTACATCTCCTACATTTGCCAAAGGGAAACTCTTTGTAACTGTAC
 TTCCAATCGAAATGACAAACGAAGCAAGTGATACCAAACCAGCAGAACTGTATCAACCT
 TGTAATCTTAGCGATATTTCACTCCCAAGCAATAGATGAACAACAAGGAAACAATTGGTA
 TACTTCCCTTGAATACCAAATTATTGATGTCAACAAGCCACTGAAAGGTTTTTCGTAATAC
 TTCTAGGTGAAATTGTTGTGGTAAATCCACTAGAAAAGCATTTAATAACAAAGCAATGG
 AACCTGTCATAATAACAGGCATAGTCCCCACAAATGAATCACGTT

ORF Predictions:

ORF #	Start	End	Direction	Length
1	1427	1711	R	95 aa

> 3864958-2 ORF translation from 1427-1711, direction R
 VDLPPQQFHLGSITKTFQWLVDINNLVFKGSIPIVSLLFYCLGVNIAKIYKVDTVSAAGLV
 SLASFVISIGSTVTKSFPLANVGDKLDQILTWN*

Description:

unknown

Assembly ID: 3865022

Assembly Length: 1386bp

> 3865022 Strep Assembly -- Assembly id#3865022

ATCGAATTTTCATTTCTATTTTCCTATTCCATTTTTTATTCAAAAAATCAAAAAGCAAACCTAG
 AAAGCTGGTCGCTGGTGGTTCAAACACTGTTTTGAGATTGTCAATAGAAGCTGACAAACC
 CTGTAATATACCTGCATATATACATACGACAAGGCGATACTACCCTAGTTTGAAGAGATT
 TTCGAAGAGTATTCATTTTTTGTCTTTTACTTATTATACCATATTCACATAAAAAAACGAA
 CATTCTTATCCTAAAAAATGCTCATTTTTTCTTAAATTATCAATCTAAATCTGGTTTATAG
 AAGGAACGATTATCCATAGCGAAGATTTTATTGGTCATCTCTCCTTTATCCACCAAAGCC
 AGAGCTGTTGACATCATCATCATGCTTGCATCCAGATTGTCAATCATATGGATAATCTCT
 GCCTCCATAATACGTGGACGGACTGGAATTTCCATATTCAAGCAAGCCGTGGTGGACTTG
 AGGATGACATGACGAAGCAAACGACTTCTTCCTTGGTATCATCGATGCCGAGTTCCATA
 ACTGTCTTGGTAATTTTCGCTATCAATGAGAGCGATATGTCCAAGAAGATTACCTCGCACT
 GTGTACTCTGTCTGGTCTGGCCCCGTCAACTCGATAACCTTAGCTAAGTCATGCAGCATA
 ATCCCCGCATAGAGCAGGCTCTTATTGAGCTGAGGATAAACTTCGCTAATAGCGTCTGCC
 AAACGTACCATGGTCGCCGTATGATAAGCCAACCCCGTTTCAAAGGCATGGTGGTTGGTC
 TTGGCGGCTGGATAGGAGTAGAATTCCTTATCATACTTGGTGTAGAGATTTTCGGACAATC
 CGTTGCCAGACAGGATTTTCAATTTTGAAAATCATTTGCGACATGTAGTCACGAATTTCC
 TTGACATCAACTGGTGACTTGACCTTGAAATCAGCTGGGTGCTTGGGTTTACCAGCTTGA
 GGCAGGCGGAGAGTAATTTGATTGACTTGAGGGGTATTGTTATAAACTTCTCGGCGTCCT
 TTCATGTGGACAACCTTACCTGCGGTAAAGGCCTCAATGTTATGAGGTTGGGCATCCCAG

AGCTTCCCATCAATCTCGCCACTATCATCTTGGAAGGTAAAGGCTAGGTAGTTTTTCCCA
 GCTCGAGTTTGCCTCAGGTCAGCTGATTTGATTAGGTAAAAGCCTTCAAATAACTCATCT
 TTTTTCATGTGACTAATCTTCATATTCTTCCTCATTTTCTTGAAAATGGAGTAGATCAAG
 CGCAGGCTCACCTTCTGACAACCTCAATGTGACGGAGCGTCCGCTCGATAGCTATGGTACG
 ACGGTTTAATAATTCGATCAATATTGCCAGAGGCATGTTGGAGATGTTTTGTGCCTTGA
 CCAGAA

ORF Predictions:

ORF #	Start	End	Direction	Length
1	279	1271	R	331 aa

> 3865022-1 ORF translation from 279-1271, direction R
 VSLRLIYSIFKKMRKNMKISHMKKDELFEFYLKISADLRQTRAGKNYLAFTFQDDSGEI
 DGKLWDAQPHNIEAFTAGKVVMKGRREVYNNTQPQVNQITLRLPQAGEPNDFKVKSP
 VDVKEIRDYMSQMIFKIENPVWQRIVRNLYTKYDKEFYSYPAAKTNHHAFETGLAYHTAT
 MVRLADAISEVYPQLNKSLLYAGIMLHDLAKVIELTGPDPQTEYTVRGNLLGHIALIDSEI
 TKTVMELGIDDTKEEVVLLRHVILKSTTACLNMEIPVRPRIMEAEIIHMDNLDASMMM
 STALALVDKGEMTNKIFAMDNRSFYKPDLD*

Description:

gi|710422 (U21636) cmp-binding-factor 1 [Staphylococcus aureus]

Assembly ID: 3865036

Assembly Length: 1167bp

> 3865036 Strep Assembly -- Assembly id#3865036
 CTCAGATTACAGAGGACAATCAACTGGTTCATTTTCGTTTCCAGTTTCAAAAAGGCTTAG
 AAAGGGAGTTCATCTATCGTGTGGAAAAAGAAAAAGTTAAGGCAGGTGTTCTCCTCTAC
 GCAGTCACCATAGCAGCCATCTTTAGTCTTTTGTGCAATTTTATTTGAACCGACAAGTC
 GCCCCTATCAAGACTATGCTTTGAATAAAGAAAAAATTGGTTGCTTTTGCTATGGCTAAA
 CGAACCAAAGATAAGGTTGAGCAAGAAAGTGGGGAACAGGTTTTTAATCTAGGTCAGGTA
 AGCTATCAAAACAAGAAAAGTGGCTTAGTGACGAGGGTTCGTACGGATAAGAGCCAATAT
 GAGTTTCTGTTTCCTTCAGTCAAAATCAAAGAAGAGAAAAGAGATAAAAAGGAAGAGGTA
 GCGACCGATTCAAGCGAAAAAGTGGAGAAGAAAAAATCAGAAGAGAAGCCTGAAAAGAAA
 GAGAATTTCCTAGTCAATTCAACTATAATGCGTTGAATCCAGAATAGTCCACTGTAGTTTC
 TAGAAAATTGCTGGAAATGGATGTTAAGCTCCAATTCATTTGTTTATATCTTATTTTCAGT
 CCACTATACTTTGTGCTAAATTAAAGATATGAAACATGATTTTAACCACAAAGCAGAAAC
 TTTTCGATTTCCCTAAAAATATCTTCCTCGCAAACCTTGGTATGTCAAGCAGCCGAGAAACA
 GATTGATCTTCTATCAGACAAAGAAATTTTAGATTTTCGGTGGTGGCACGGGTCTATTAGC
 CTTGCCCCCTAACCCCTAGCCAAGCAGGCTAAGTCAGTCACTCTTGTAACATTTCTGAGA
 AAATGTTGGAGCAAGCTCGTTTGAAAGTGGAGCAGCAAGCAATCAAGAATATCCAGTTTT
 TGGAGCAAGATTTACCGAAAAATCCCTTGGAGAAAGAGTTTGATTGCCTTGCTGTAGTC

GGGTTCTTCATCATATGCCTGATTTGGATGCGGCTCTCTCACTGTTTCATCAACATTTGA
 AGGAAGATGGGAACTCATCATTTGCTGATTTTACCAAGACAGAAGCTAATCATCATGGAT
 TTGATTTAGCTGAACTGGAAAACAAGCTAATTGAGCATGGGTTTTTCATCTGTGCATAGT
 CAGATNCTCTATAGCGCTGAAGANCTG

ORF Predictions:

ORF #	Start	End	Direction	Length
1	79	492	F	138 aa

> 3865036-1 ORF translation from 79-492, direction F
 VWKKKKVKAGVLLYAVTIAAIFSLLLQFYLNQVAHYQDYALNKEKLVAFAFAMAKRTKDKV
 EQESGEQVFNLGQVSYQNKKKTGLVTRVRTDKSQYEFLLFPSVKIKKEEKRDKKKEEVATDSSE
 KVEKKKSEEKPEKKENS*

Description:

unknown

Assembly ID: 3865054

Assembly Length: 916bp

> 3865054 Strep Assembly -- Assembly id#3865054

TCTCCCAACATATAATTTCCGTTTTTCCAATCCCCCAGCTGTCATACAGTCTGTGATAAGA
 GCGATGTTTTCTGTTTCCTTTTTTGTGTTGATAAGAATTTTCGCAAGCCTTTGGATCTACGTGG
 TGACCATCACAGATCAACTCTGCATAGGTATGTGGCAATTGGTACATGGCTCCAACCATA
 CCCAATTCACGGTGAGTCAACCCACGCATTCCATTGTAGGCATGCACCCAAACACTCGCT
 CCAGCATCGACTGCTTTTTTGGCTTCATCAAAAGTCGCGTTTGAATGTCCAAGAGCAACC
 GTCACACCTTCGCCCCGTAAGTGTACGAACAAAGTCTTCCACCCCATCACGTTCTGGTGCA
 ATCGAATTTTATTAAGCAAGCCATTTGCCGCTTTTTTGCCAAGAATGAACTCCTCAACAC
 CCGGGTCTCTCATATAAGTTGGATTTTGTGCCCCCTTAAAAGTTTCTGTGAAATATGGAC
 CTTCATAATAAATCCCACGAATCTTAGCACCTGTTGCTTCTTTATAATGGTTTCCAAGAT
 TTTCAGTGACTGCAAGCAATTGCTCATAAGTGGCTGTTAAAGTTGTGGGTAAGAACTGG
 TAACACCGGTACTAAGAAGTCCTTCACTCATAGTATGCAATGTACCTTCAATGTTGTTGT
 CCATCACATCTACACCTGCATATCCATGAATATGAGTATCCACAAGACCTGGGGCAATGC
 TATAACCTGTATAGTCAATCACCTCAGCCCCCTCAGGAATCTGCTCTACATGTTTCCCAA
 ACTTGCCGTCCACAAGTTCCAAGTAACCACCTCGACAAATCCGTGTGGGTAGAAAACTG
 ATCCGCTTTAATATAGTTAGGCATAATGTTAACCTCCTTAAAAGATTGATTCTACAATTT
 ATTATGTCAATTTCGAT

ORF Predictions:

ORF #	Start	End	Direction	Length
1	302	793	R	164 aa

> 3865054-1 ORF translation from 302-793, direction R
 VDGKFGKHVEQIPEGAEVIDYTGYSIAPGLVDTHIHGYAGVDVMDNNIEGTLHTMSEGLL
 STGVTSFLPTTLTATYEQLLAVTENLGNHYKEATGAKIRGIYYEGPYFTETFKGAQNPTY
 MRDPGVVEEFHSWQKAANGLLNKIRLHQNMVGWKTFLFVQLRAKV*

Description:

N-acetylglucosamine-6-phosphate deacetylase (nagA) homolog -
 Haemophilus influenzae (strain Rd KW20)

Assembly ID: 3865102

Assembly Length: 786bp

> 3865102 Strep Assembly -- Assembly id#3865102
 CTGGATTAAAACGAGGCAGTTTCAGACTAATATCCAAGTCGTAAGAAATGCCTGAAATAA
 GCTTTTCTAAATTGTCCAAAGCTTGCGGGAAAACGCTCTTGGAATAGTTTCTCTAAAGAA
 CTTGCTGATATAAAGACATCTTGTCTCGAACGCAAGGGAACCTTCTCTGAGCGGTAGATTT
 TCTTTAATCGCTGTTAAAACTTGAAGAACTTCTCTATCCCTGCTTTCAAAGCGTTGACC
 CGATAAAGAGGTAAGATAGGATGATGAAATTCGCTTGCTAGTGTCTTGATAAACCCCT
 ATATAGTAATCACAGCCTAGTTCTAACGACTCAACTCTATCAAAATAAGGCACAATGACC
 GCGATATCCTCCAGGTACTGGGACAGGACTGACCAAGTTTCTCCCCCTGCATCTTGGCT
 GTCGAAAGCTTCATCAACTGCTGATAGCCACACTAGATAGAGCTAAAAAGCGCAAATTC
 ACTTCCTGATCATCTACAAACACTGTCATTTCAAGCCCTAGCAAAGGATGAATGCCGTAT
 TTTTTTGTAATCTCTAGAAAGTCGAAAGCGCCATAAAGATTGTCAATATCCATCATAGCC
 AAATGAGTGTAGCCGTATTCTTTAGCTGCTCTCACATACTTTTCGATCGAAATGACGCTT
 TCCATAAAACTATAGACTGTTTTGTATCTAGTTGTGCGATCAATTTACACTTCTCCTCT
 ATCCTTCTCACTATATTATACCATTTTCACCTATAAATGGCTTCTCTTGAGAAAAATTC
 GATCAG

ORF Predictions:

ORF #	Start	End	Direction	Length
1	27	731	R	235 aa

> 3865102-1 ORF translation from 27-231, direction R
 VRRIEEKCKLIAQLDTKTVYSFMESVISIEKYVRAAKEYGYTHLAMMDIDNLYGAFDFLE
 ITKKYGIHPLLGLEMTVFVDDQEVNLRFLALSSVGYQQLMKLSTAKMQGEKTWSVLSQYL
 EDIAVIVPYFDRVESLELGCDDYIGVYPETLASEFHHPILPLYRVNAFESRDREVLQVLT
 AIKENLPLREVPLRSRQDVFISSASLEKLFQERFPASFGQFRKAYFRHFLRLGY*

Description:

unknown

Assembly ID: 3865156

Assembly Length: 1213bp

> 3865156 Strep Assembly -- Assembly id#3865156

CACTTTCAGCTTCTTCTCTTTTTGAACGGTTATAAACACGAATCAGATTCCCTATTTCTT
GCGATTTATGTGATTCCTTATTTTCCAATCTAAAGTATAGTGAAATGAAATAAACATGC
GCAAATCGATTAAGGAATTTAATCTAATTTCTAACAATGTCTTAGAAATCAAAGTGACT
ATTTTAACTTCAATGCACTAAACATCTAATACTCAATAAAAAATCAAAGAGCAAAGTAGGA
AACTAGCCGCAGGTGGCTCAAAACACTGTTTTGAGGTGTGTAGATGAAACTGACGAAGTCA
GTAACCATACATACGGCAAGGCGACGCTGACGTGGTTTGAAGAGATTTTTCGAAGAGTAGC
AAAATGGAAAAAGGAGTGAGTGAAGCACATCGCCTCCCCACTCCTTTTTCTGTTTTTAGG
CTGTTTTTTTCAACCTTCAAGATTTTTTACATCATAGCTACCAACAGGCGTTTTCAATGGTTG
CTGTATCACCTGTTTTCTTGCCAATCAAGGCCTGCCCAATTGGGCTTTCATTTGAAACCT
TACCTGCAAAGGCATCCGCACCAGCTGAACCTACGATAATATAAACTTCTTCTTCGTCCT
CACCAATTTCTTGGATGGTGACTGTTTTACCAATCGCTACTTCGTCCTGGGCAACTGCGT
CGCTATTGACGATTTTACGCATAGCGGATTTTTGTTTTCTAAGCTAGAGATTTGTCCTTCGA
CAAAGGCTTGTTTCATCCTTAGCTGCTTCGTAATCACTGTTTTCTGAAAGGTCACCGTATG
AACGGGCAATCTTAATGCGTTCTACCACCTTCTGGTTCGACGAAACCAATTTCAATTCTTCT
AATTCTTTTTCAAGTTTTTTCCTTTTCTCAAGGGTCATAGGATATGTTTTTCTGCCATT
TTTCTCAACTTTCTTCTGATAATATTTTCTAAAGAAAATTATGTGAAGTATCACATAATT
TTAGTTTGTGTTAGTTTAATTTGCTGTTGACATGTTTCAGCGACATTGCGGTCTGGTCTTC
TTGATTGTTAGCATAGTAAACCTTGCCCTTCTGTGACATCTGCTACAAAGTAAAAGTTATC
GCTCTTAGTTTGATTGATGCTTGACTCAATCCGCATCCAAGACTTGGACTATCGACTGGA
CCAGGCATGAGACCTACATTTTTTATAAACATTATAAGGTGAATCAATGTTGGTATCAATC
GCAACATCCTCAG

ORF Predictions:

ORF #	Start	End	Direction	Length
1	416	808	R	131 aa

> 3865156-1 ORF translation from 416-808, direction R

VVERIKIARSYGDLSENSEYEAAKDEQAFVEGQISSLETKIRYAEIVNSDAVAQDEVAIG
KTVTIQEIGEDEEEVYIIIVGSAGADAFAGKVS NESPIGQALIGKKTGDTATIETPVGSYD
VKILKVEKTA*

Description:

TRANSCRIPTION ELONGATION FACTOR GRE A (TRANSCRIPT CLEAVAGE FACTOR
GRE A). - ESCHE RICHIA COLI.

Assembly ID: 3865160

Assembly Length: 1173bp

> 3865160 Strep Assembly -- Assembly id#3865160

TGCGGCTGAGTTGGGAATTCCTATCGTTAATAAGCGTGTATCGGTGACACCTATTTCTCT
 GATTGGGGCAGCGACAGATGCGACGGACTACTGGTTCTGGCAAAAGCGCTTGATAAGGCT
 GCGAAAGAGATTGGTGTGGACTTTATTGGTGGTCTTTCTGCCTTAGAACAAAAAGGTTAT
 CAAAAGGGAGATGAGATTCTCATCAATTCCATTCTCGCGCTTTGACTGAGACGGATAAG
 GTCTGCTCGTCAGTCAATATCGGGCTCAACCAAGTCTGGTATTAATATGACGGCTGTGGCA
 GATATGGGACGAATTTATCAAGGAAACGGCAAATCTTTCAGATATGGGAGCGGCCAAGTT
 GGTGTATTTCGCTAATGCTGTTGAGGACAATCCATTTATGGCGGGTGCCTTTCATGGTGT
 TGGGGAAGCAGATGTTATCATCAATGTCGGAGTTTCTGGTCCTGGTGTGGTGAAACGTGC
 TTTGGAAAAAGTTCGTGGACAGAGCTTTGATGTTAGTAACCCGAAAACCAGTTAAGAAAA
 CTGCCTTTTAAATCACTCCGTATCCGGTCCAATTGGTTTGGTCAAATGCCCAGTGAGAG
 ACTGGGTGTGGAGTTTGGTATTGTGGACTTGAGTTTGGCACCAACCCCTGCGGTTGGAGA
 CTCTGTGGCACGTGTCCTTGAGGAAATGGGGCTAGAAACAGTTGGCACGCATGGAACGAC
 AGCTGCCTTGGCCCTCTTGAACGACCAAGTTAAAAAGGGTGGAGTGATGGCCTGTAACCA
 GGTCGGTGGTCTATCTGGTGCCTTTATCCCTGTTTCTGAGGATGAAGGAATGATTGCTGC
 AGTGCAAAATGGCTCTCTTAATTTAGAAAACTAGAAGCTATGACGGCTATCTGTTCTTG
 TTGGATTGGATATGATTGCCATCCCAAGATACGCCTGCTGAAACTATTGCGGCTATGA
 TTGCGGATGAAGCAGCAATCGGTGTTATCAACATGAAAACAACAGCTGTTTCGTATCATTC
 CCAAAGGAAGAGAAGGCGATATGATTGAGTTTGGTGGTCTATTAGGAAGTGCACCCGTTA
 TGAAGGTTAATGGGGCTTCGTCTGTCGACTTCATCTCTCGCGGTGGACAAATCCCAGCAC
 CAATTCATAGTTTAAATAATTAAGAAAATAGGA

ORF Predictions:

ORF #	Start	End	Direction	Length
1	136	375	F	80 aa

> 3865160-1 ORF translation from 136-375, direction F
 VDFIGGLSALEQKGYQKGDEILINSIPRALTETDKVCSSVNIGSTKSGINMTAVADMGRI
 YQNGNKSFRYGSQVGCIR*

Description:

unknown

Assembly ID: 3865172

Assembly Length: 1209bp

> 3865172 Strep Assembly -- Assembly id#3865172

TCGGAATCTGAGCTAGTGTAGCTTCCTTAATCTTATCTGATAAGATAGCTGTCATATCAG
 ACTCAATCATTTCTGGAGCAATCAACATTGACTCGTATATTCCGACTAGCGACCTCGCG
 TGCCACAGACTTGGTAAAGCCAATCAAGCCAGCCTTAGAAGCAGCATAGTTAGCTTGACC
 AATATTCCTCATCAAACCAACAACACTAGACATATTAATGATAGCACCTTCTCTGGCTTT
 CATCATCGGTTTCAAGACTGATTGTGTCATATTAAAGGCACCAGTCAGATTGACCTTGAG
 CACTTTTTCAAAATCTGCTTCTGTCATCTTGAGCATAAGAGTATCTTGGGTAATCCCTGC

ATTGTTGACCAAAACATCTACTGAACCCAGTTCTGCAATAGCTTGATCAATCATACGCTT
 AGCGTCTGCAAAATCTGATACATCTCCTGAAATGGGAACCACCTTGATACCATAGTTTGA
 AAACCTCAGCGAGCAATTCTTCTGAGATTGCCCCACGACTGTTTAAGACAATGTTGGCTCC
 TGCTTGAGCAAACCTTGTGGGCGATGGCAAGACCAATTCACGACTCGAACCTGTAATAAA
 GATATTTTTTATGTTCTAGTTTCATTTTTTTTCCCTTTCAAAACCTTCTACTTATTTTAGTCTA
 TTTTTCTAAAAGTGCTACTAAACTCGCTTGATCTTCCACATGAGCTAAGTGAGCAGTTTG
 ATCAATTTTTTTTAAACAAAACCTGACAAGACTTTCCTCCGCTCCAATCTCGAATAAAGTTGC
 TTATGCCTGCTTCTTGATGACCCCAATACTTTCATAGAAACGAACGGGTTCCCTTGACCT
 GACGCGTCAAGAGCTGAGCAATGTCCTCTTTTTTGCATCACAGCAGCTTCTGTATTGCCGA
 CTAGGGGACAAGTAAATCTGAAAAACTTACCTGAGCTAGAGTTTCAGCTAGTTTCTGGC
 TAGCAGGCTCAAGGAGAGCGGTGTGAAAGGGACCTGACACCTTAAGAGGAATCAAGCGTT
 TGGCACCTGCTTCTTGCAAAAGTTCAACCGCTCGATCAACTGCAACCACTTCTCCAGCAA
 TGACGATTTGTGCAGGTGTGTTATAGTTGGCTGGAGTAACCACTCCAAGTTCCAGAAGCT
 TTTTGACAGGCTTCTTCAATGACCTCTACTGGCGTATTGAGAACTGCTACCATCTTGCCA
 AGTTCAGCA

ORF Predictions:

ORF #	Start	End	Direction	Length
1	731	1123	R	131 aa

> 3865172-2 ORF translation from 731-1123, direction R

VVTPANYNTPAQIIVIAGEVVAVDRAVELLQEAGAKRLIPLKVS GPFHTALLEPASQKLAE
 TLAQVSFSDFTCPLVGNTTEAAVMQKEDIAQLLTRQVKEPVRFYESIGVMQEAGISNFIRD
 WTGESLVRFC*

Description:

malonyl coenzyme A-acyl carrier protein transacylase (fabD)
 homolog - Haemophilus influenzae (strain Rd KW20)

Assembly ID: 3865228

Assembly Length: 813bp

> 3865228 Strep Assembly -- Assembly id#3865228

ATGACACGTCTGTTCTCTCAAGCAGAAATGGCAGAGTAACAAGCTCGATATTGAGGTAGC
 CGATAAAGAATTGGCTGAATTTGAAGCTCAGATTAAACAGGAAGTGGAAGCTCCAACTTG
 TAGTGAGTCCTCAGGTTGAAGAAGAGCCTCAGCTCATCCAGTTGGCCCAATGTATGAAGA
 ACCAGAAGTAAATCCAGTGCATCCGACAGGTCCAACACCAGCTACAGAACTGTTGATTC
 AATACCGGGATTTGAAGCACCGCAAGAATCTGTTACAATTTTATAAGAAATATTCTGAGA
 ACAATATCTTATCCTTATATTTCCAGCGAGCAGGAAATGGTGTGAGTCCTGCATTCCCTA
 TCGATAAGATTATCCTCTCAAACCTATCAAGTCTGAATCTAGTAAGATTTGACGTTCCCA
 CGTTACGGGATAAGAGAGAGAAAGACTAAATCTTTTTTCCGAATAAAGGTGGTACCACGAT
 TTTCGTCCTTTTTTGAAGTCGTGGTTTTTAATTTGTTATTATTATAAAGGAGATACCAT

GAAACTCAAAGACACCCTTAATCTTGGGAAAACCTGAATTCCCAATGCGTGCAGGCCTTCC
TACCAAAGAGCCAGTTTGGCAAAGGAATGGGAAGATGCAAACTTTATCAACGTCGTCA
AGAATTGAACCAAGGAAAACCTCATTTACCTTGCATGATGGCCCTCCATACGCTAACGG
AAATATCCACGTTGGACATGCTATGAACAAGATTTCAAAAGATATCATTTGTTTCGTTCTAA
GTCTATGTCAGGATTTTACGCGCCATTTATTCC

ORF Predictions:

ORF #	Start	End	Direction	Length
1	197	286	F	30 aa

> 3865228-1 ORF translation from 197-286, direction F
VHPTGPTPATETVDSIPGFEAPQESVTIL*

Description:

unknown

Assembly ID: 3865230

Assembly Length: 953bp

> 3865230 Strep Assembly -- Assembly id#3865230

ATCGAATTATTTTGAACAAGGTGGATCAGCTATTTTGGCCTTGATTAGTATTTTACTCT
TTAAATACACTTGAAGGTCGATTCTAATCTCGCTAATCCTTTTAAATCCAGAATAAGGGA
AATATGTTATACTTGTTTTTAAGAAAAAAGTTTCATTGAATTGGTTTTGAGGAGTTAGAA
ATGAAAGTATTAGTGACAGGTTTTTGAGCCCTTTTGAGGCCATTAAAGGTTTACCAGCTGA
AATCCATGGTGCTGAGGTCCGTTGGCTAGAGGTGCCGACAGTTTTTCACAAATCTGCTCA
AGTATTGGAAGAAGAGATGAATCGTTATCAACCTGACTTTGTCCTTTGTATTGGGCAAGC
TGGTGGAAGAACTAGTTTGACACCTGAACGAGTGGCCATTAAATCAAGACGATGCACGTAC
TTCTGATAACGAAGATAATCAACCGATTGACCGTCCCATTTCGCCCAGATGGTGCTTCGGC
CTACTTTAGTAGTTTGCCGATTAAAGCGATGGTTCAAGCTATAAAAAAGAAGGATTACCG
GCCTCTGTTTCCAATACGGCAGGGACTTTTGTCTGCAGCCATTTGATGTATCAGGCTCTC
TATTTGGTAGAAAAGAAATTCCCATATGTTAAGGCAGGTTTTATGCATATTCCTTATATG
ATGGAACAGGTGGTGAACAGACCGACTACTCCAACCTATGAGTTTAGTGATATTCGGCGA
GGGATAGAAGCAGCAATCGGCGCTATGATAGAACATGGAGATCAGGAACCTCAAGTTGGTA
GGCGGAGAAATTCATTGATAGAAAAAAGCTTGAGGGGAAAACCTTCAAGCTTTTGGACGT
TTTCGAGCCAATACTGCTCGGTAAAACATAATTTTAGTGCAATTGGATATAAGGTAGGAGT
GAAAACTAGCAATGCCAAAGGTAATCCAATTGAGGAAGTACCAAGGAAGAAG

ORF Predictions:

ORF #	Start	End	Direction	Length
1	272	586	F	105 aa

> 3865230-1 ORF translation from 272-586, direction F
 VPTVFHKSQAQVLEEEMNRYQPDFVLCIGQAGGRTSLTPERVAINQDDARTSDNEDNQPID
 RPIRPDGASAYFSSLPIKAMVQAIKKKDYRPLFPIRQGLLSAAI*

Description:

PYRROLIDONE-CARBOXYLATE PEPTIDASE (EC 3.4.19.3) (5-OXOPROLYL-
 PEPTIDASE). - STR EPTOCOCCUS PYOGENES.

Assembly ID: 3865378

Assembly Length: 1060bp

> 3865378 Strep Assembly -- Assembly id#3865378

CTACTTGAAACAGAACTGAAATTATACCCACTACCTCCCTGATTATCTTCAATGCTTACG
 TCTAAATAAACTTCCCCACTATTATTTAGCTTAGCAACAACTGTTATAGTAAAATAACAT
 AAAATTACATAAAATAGATTAGGGAAATCAAAGCAACTTCTAGGAATGTTTTAGCAGTCA
 CAGTGTACTTTCCAGCATCAAGCCACTATAACTCTGCACATAAAAATGGAGAAGATGGC
 CATCCTCTTCTCCAAATATTAACCTTCTTTACAAACCAACTATAGTTGACAAAGAACCTAA
 AATCAATTGATAACACGAGGTCAGGTCGGTCAACTCTTTCAACTGAAGCCCTGTCAACTC
 TTCCCATTTATCAATCTTGTATTGGAGAGAATTGCGGTGCAGATAGAGTTGCTGGGCTGT
 TTAAGTGAGAACAGCACTATTTTCCCAAAGAGAGAGAATGATTTCTGAATCTGATCTTG
 ATCCAAAATCATCTGGTGTAGACATTCCTTGATTGGCTTCAAGTCCACGAGTCTTTCTCC
 CAGACTCCAAAGATAGAGCTGAGAAAAAGTATGAACACCTTGGTGACCCTGACGCCACCA
 TGTCTTGAACAAATCCCGCTCAGCTTTGATTAAGTCTGATAGGGCTTGATGTCCCGTCTG
 AGACCAAACCTGACCCAACATGATAGAAAGACGAAGTCCAAAGTCATACTCAACCGCTTC
 AATCGTATCACTTAAAATATCTCTTACAGAAAGTGTATTTGTCTTGTTGAAGCACGAAAAC
 ATAATCCTGAGATCCGACCTGTAGCACTGTCTGACAATTCGGAAAAAGAGTCCGCATCAT
 ATCTAGCCAAGAAGCCAGATTTTCTCTGCTGAAAATAAGAAAGATGGCAATAAACCAACTG
 AATCTTTTTTAAAAACTTGCGGTGCCTGTCCCTTGCCCTCAACCAGATAGGAATACCAAGG
 GTTTAGCGAACGAACCTGCTCCTGCTGGGTCAAAGGGCAACCAACTGCTTTTCACGCTC
 GCTGAGCCCAGCTTCCTCCAGCAAAATCCACTGCTGAGAG

ORF Predictions:

ORF #	Start	End	Direction	Length
1	421	807	R	129 aa

> 3865378-1 ORF translation from 421-807, direction R
 VLQVGSQDYVFLVQQDKYTSVRDILSDTIEAVEYDFGLRLSIMLGQVWSQTGHQALSDLI
 KAERDLFKTWWRQGHQGVHTFSQLYLWSLGERLVLDKPIKECLHQMILDQDQIQEIILSL
 WENSAVLT*

Description:

unknown

Assembly ID: 3865470
 Assembly Length: 895bp

```
> 3865470 Strep Assembly -- Assembly id#3865470
ATTTTAGACTTTGATGACAATCCTCAGGCGGTTATCATGCCCAATCACGAGGGGCTGGAA
TTGCAGTTGCCAAAGAAGTGTGTTTATGCATTTTTAGGTGAGGAGATCTGACCGCTATGC
AAGGGAAGTAGGGGCGGATTGTGTCGGCGAATTCGTTTCTGCTACCAAGACCTATCCAGT
CTCTTTCATCAACTACAAGGGTGAGGAGGTCTGTCTGGATCAGGCTCCTGCTGGCTCCGC
TCCAGCAGCCCAGTTTATGGATGGGTTGATTGGCTATGGTGTGGAGCAGCTTATCTCTAC
TGGGACCTGTGGTGTCTTAGCTGATATAGAGGAAAATGCCTTCTAGTCCCTGTTTCGCGC
TTTGCGAGATGAGGGAGCCAGTTACCACCTATGTGGCACCTTGTTCGTTATATGGAAATGCA
GCCAGAGGCTATTGCTGCTATTGAGGAAGTTTTGGAAGACAGAGGGATTCTTATGAAGA
AGTCATGACCTGGACGACAGACGGTTTTTACCGAGAAACGGCTGAAAAGGTGGCTTATCG
TAAGGAAGAAGGCTGTGCTGTTGTGGAGATGGAGTGTTCTGCTCTTGCGGCAGTAGCTCA
ATTGCGTGGGGTTCTCTGGGGTGAATTGTTGTTTACAGCAAATTCTCTAGCGGACTTGGA
CCAGTACAACAGTCGTGACTGGGGCTCGGAACCTTTTAATAAGGCGCTAAAACTGAGTTT
AGCAAGTGTCCACCACCTTTAGTTGTACTGGCAAAGGATTTGTTTTATCATAAAATGTCT
AGCTCATACTTTTCAAAAATATGTTTAAACGAAGTCACCTTCCTCTTGTCCTAAGCATGT
TTGAAGTTGGGAAAAATCTTTAAATCAGAAAAACGTATCATATCAGGTTGATGA
```

ORF Predictions:

ORF #	Start	End	Direction	Length
1	98	742	F	215 aa

```
> 3865470-1 ORF translation from 98-742, direction F
VRRSDRYAREVGADCVGEFVSATKTPVVSFINYKGEEVCLDQAPAGSAPAAQFMDGLIGY
GVEQLISTGTCGVLADIEENAFVLPVRLRDEGASYHYVAPCRYMEMQPEAIAAIEEVLE
DRGIPYEEVMTWTTDGFYRETAEKVAYRKEEGCAVMECSALAAVAQLRGVLWGELLFT
ANSLADLDQYNSRDWGSEPFNKALKLSLASVHHL*
```

Description:
 unknown

Assembly ID: 3865632
 Assembly Length: 645bp

```
> 3865632 Strep Assembly -- Assembly id#3865632
AGGGCTGTCAAGCTTGGTTAGAACGTTTAGAAAAGGAGAGTTAAGGTGGAAAATCTTACG
AATTTTTACGAAAAGTATCGTGTCTATCTGACTCGTCCACGTTTAGAGCTTTTGGCAGTA
GTTACCATTGTTTTANGNGCTGTACTCGTCTTTTTTCTAAATATTCAGGAAAAGGTGTC
TTAAAACTCGATAATGGAACGATTGTTTATGATGGCAGTCTTGTCCGTGGTAAATGAAT
```

GGCCAAGGTACCATTACCTTCCAAAATGGAGACCAATATACAGGTGGCTTCAACAATGGA
 GCCTTCAACGGAAAAGGTACCTTTCAATCTAAAGAAGGCTGGACCTACGAAGGTGATTTT
 GTAAATGGTCAGGCTGAAGGAAAAGGGAACTAACAACAGAACAAGAAGTCGTTTATGAA
 GGAACTTTAAACAAGGCGTTTTTCAACAAAAATAAAGCCTCCTTATCAAAGGAGGTATT
 ATTAGAATTACAAGGTAAGCGTTTACCTGTAAATCCCTTTCTTTCCAAATCCCTCTTCCA
 AGCAAGTTTGTGAAATAAAAAATATTTGAAATAAATTTACAAACTTCAAAGATAAAACC
 TGATAAGAAAAGAAAATGAGAAAAGTTTCGCAAGAGTTTAAAAAT

ORF Predictions:

ORF #	Start	End	Direction	Length
1	46	456	F	137 aa

> 3865632-1 ORF translation from 46-456, direction F
 VENLTNFYEKYRVYLTRPRLELLAVVTIVLXAVLVFFLNIPGKGVLKLDNGTIVYDGSLV
 RGKMNQGTITFQNGDQYTGGFNNGAFNGKGTFFQSKEGWTYEGDFVNGQAEGKGKLTTEQ
 EVVYEGTFKQGVFQQK*

Description:

unknown

Assembly ID: 3865710

Assembly Length: 572bp

> 3865710 Strep Assembly -- Assembly id#3865710
 GAGATCTGTCTTGACACCAAAGTGTGGAGTACGCCAGCTAATTCAACGGCGATATAACC
 AGCGCCTAGAATCGCAATTGACTCTGGAAGTTCTTCCCAGGCAAATACATCATCAGAAGA
 GCCACCTAGCTCAGCACCAGGAATATTAGGAATACTTGGATGGGCACCTGTAGCAATCAC
 GATATGTCTAGCACGAATCAGTTCACCATTTACGCTTACAGTATGAGAATCTACAAATTC
 AGCATGACCTTCAATCAAGTCTACACCGTTGCGTTTAAAACTACCATCATAGAGAAGAAC
 GAGCGCGATCAATGTAGGCTTCACGATTGCGACGTAGGGTTGCAAAGTTAAAGTTAAGAT
 CAGTAGTCTCAAAGCCGTAGTCTCCTCCAAATTGATGGAAAGTCTCAGCGATTTGCGCCC
 CGCTACCACATGATTCTTTTAGGAACACAACCGACGTTGACACAGGTTCCACCTAATTTT
 TTTTCTCAATAACGGCTGCTTTGGCTCCATGTTCCCAGCACGGTTCATGGTAGCGATCC
 TCCGCTACCTCCACGATAGCAATGATATCATA

ORF Predictions:

ORF #	Start	End	Direction	Length
1	287	448	R	54 aa

> 3865710-1 ORF translation from 287-448, direction R
 VFLKESCGSGAQIAETFHQFGGDYGFETDNLNFNFATLRRNREAYIDRARSSL*

Description:

glutathione reductase (NADPH) (EC 1.6.4.2) - *Streptococcus thermophilus*

Provided in Table 2 is information on the direction of the ORF (forward or reverse) for each polynucleotide in Table 1. Also listed for each ORF is its start and stop codon positions (refer to the columns containing nucleotide code labeled "Start" and "Stop"). The triplet codon sequence for each start and stop codon is also shown. These codons may be shown in the sense orientation or antisense orientation, such as GTG and CAC, respectively, for start codons. The "Length" column discloses the length of each polynucleotide assembly. The direction of translation on the polynucleotide depicted is denoted by and "Forward" for forward or and "Reverse" for reverse (or being on the opposite strand from the one depicted). As indicated above, the "Assembly ID" number is a unique identifier assigned to each ORF of Table 1 and allows a correlation between the data in Tables 1 and 2.

TABLE 2

Assembly ID	Start	Stop	Start	Stop	Length	Direction
3049156	~CAC	TCA~	236	385	50	Reverse
3049862	GTG	TGA	383	526	48	Forward
3112810	~CAC	TTA~	601	804	68	Reverse
3112866	~CAC	TTA~	220	513	98	Reverse
3113664	GTG	TAA	165	392	76	Forward

Assembly ID	Start	Stop	Start	Stop	Length	Direction
3113716	~CAC	TTA~	94	291	66	Reverse
3174176	GTG	TAA	139	543	135	Forward
3174186	GTG	TAG	83	283	67	Forward
3174374	GTG	TGA	154	294	47	Forward
3174972	~CAC	TTA~	169	678	170	Reverse
3175138	~CAC	TCA~	79	945	289	Reverse
3175860	GTG	TAA	51	251	67	Forward
3175918	GTG	TGA	212	535	108	Forward
3811220	~CAC	CTA~	316	873	186	Reverse
3811436	~CAC	TTA~	1164	1511	116	Reverse
3811984	GTG	TGA	134	454	107	Forward
3857228	~CAC	TCA~	1141	1356	72	Reverse
3857842	GTG	TAA	45	341	99	Forward
3857996	GTG	TAA	58	456	133	Forward
3858236	~CAC	CTA~	1	261	87	Reverse
3858264	~CAC	TCA~	439	1365	309	Reverse
3858610	~CAC	TTA~	374	949	192	Reverse
3858716	~CAC	CTA~	238	402	55	Reverse
3859124	~CAC	CTA~	73	453	127	Reverse
3859244	~CAC	TTA~	310	462	51	Reverse
3859250	~CAC	CTA~	244	402	53	Reverse
3859588	~CAC	TTA~	102	443	114	Reverse
3859774	~CAC	CTA~	9	131	41	Reverse
3860140	GTG	TAA	302	511	70	Forward
3860140	GTG	TAA	605	856	84	Forward
3860206	~CAC	TTA~	898	1056	53	Reverse
3860270	GTG	TAG	346	966	207	Forward
3860438	GTG	TAG	1	276	92	Forward
3860438	GTG	TGA	460	1128	223	Forward
3860544	GTG	TAA	222	689	156	Forward
3860558	~CAC	TTA~	717	1376	220	Reverse
3860568	GTG	TAA	1040	1291	84	Forward
3860582	GTG	TGA	356	1027	224	Forward
3860724	GTG	TGA	139	498	120	Forward

Assembly ID	Start	Stop	Start	Stop	Length	Direction
3860724	GTG	TGA	686	1024	113	Forward
3860858	GTG	TAG	610	807	66	Forward
3860890	GTG	TAG	397	486	30	Forward
3860952	~CAC	TTA~	449	715	89	Reverse
3860962	~CAC	TTA~	152	646	165	Reverse
3861268	~CAC	TTA~	457	645	63	Reverse
3861270	~CAC	TTA~	627	824	66	Reverse
3861288	~CAC	CTA~	357	572	72	Reverse
3861306	GTG	TAA	717	1208	164	Forward
3861306	GTG	TAA	1201	1410	70	Forward
3861334	GTG	TAA	76	975	300	Forward
3864148	GTG	TAG	212	940	243	Forward
3864148	GTG	TAA	1202	1753	184	Forward
3864148	GTG	TAA	2750	3037	96	Forward
3864172	GTG	TAG	311	862	184	Forward
3864180	~CAC	TTA~	930	1616	229	Reverse
3864184	GTG	TGA	197	670	158	Forward
3864184	GTG	TAA	612	1304	231	Forward
3864194	~CAC	CTA~	1084	1380	99	Reverse
3864338	GTG	TGA	552	1100	183	Forward
3864360	GTG	TAA	47	1078	344	Forward
3864388	GTG	TGA	1239	1586	116	Forward
3864406	~CAC	TTA~	263	958	232	Reverse
3864452	~CAC	TCA~	1079	1201	41	Reverse
3864458	GTG	TAA	797	1105	103	Forward
3864458	GTG	TGA	1179	1391	71	Forward
3864474	~CAC	CTA~	68	247	60	Reverse
3864474	~CAC	TTA~	644	1528	295	Reverse
3864510	~CAC	TTA~	1164	1640	159	Reverse
3864526	~CAC	TTA~	845	1660	272	Reverse
3864548	GTG	TGA	687	1055	123	Forward
3864548	GTG	TAA	979	1932	318	Forward
3864582	~CAC	TTA~	317	550	78	Reverse
3864604	~CAC	CTA~	1	141	47	Reverse
3864604	~CAC	CTA~	1513	1803	97	Reverse
3864610	GTG	TAA	427	1305	293	Forward
3864716	GTG	TAA	57	272	72	Forward
3864718	GTG	TGA	77	1474	466	Forward
3864802	~CAC	TTA~	92	550	153	Reverse

Assembly ID	Start	Stop	Start	Stop	Length	Direction
3864854	~CAC	CTA~	324	548	75	Reverse
3864862	~CAC	CTA~	431	1003	191	Reverse
3864888	~CAC	TTA~	10	657	216	Reverse
3864898	GTG	TAA	130	1029	300	Forward
3864938	GTG	TGA	883	1326	148	Forward
3864956	GTG	TAA	1030	1251	74	Forward
3864958	~CAC	TCA~	1427	1711	95	Reverse
3865022	~CAC	TCA~	279	1271	331	Reverse
3865036	GTG	TAG	79	492	138	Forward
3865054	~CAC	TCA~	302	793	164	Reverse
3865102	~CAC	CTA~	27	731	235	Reverse
3865156	~CAC	TTA~	416	808	131	Reverse
3865160	GTG	TAA	136	375	80	Forward
3865172	~CAC	TTA~	731	1123	131	Reverse
3865228	GTG	TAA	197	286	30	Forward
3865230	GTG	TGA	272	586	105	Forward
3865378	~CAC	TTA~	421	807	129	Reverse
3865470	GTG	TAG	98	742	215	Forward
3865632	GTG	TAA	46	456	137	Forward
3865710	~CAC	TCA~	287	448	54	Reverse

EXAMPLES

The examples below are carried out using standard techniques, which are well known and routine to those of skill in the art, except where otherwise described in detail. The examples are illustrative, but do not limit the invention.

Example 1

Isolation of DNA coding for a virulence gene in *Streptococcus pneumoniae*

As mentioned above each of the DNAs disclosed herein by virtue of the fact that it includes an intact open reading frame is useful to a greater or lesser extent as a screen for identifying antimicrobial compounds. A useful approach for selecting the preferred DNA sequences for screen development is evaluation by insertion-duplication mutagenesis. This system disclosed by Morrison et al., *J. Bacteriol.* 159:870 (1984), is applied as follows.

Briefly, random fragments of *Streptococcus pneumoniae*, strain 0100993 DNA are generated enzymatically (by restriction endonuclease digestion) or physically (by sonication based shearing) followed by gel fractionation and end repair employing T4 DNA

polymerase. It is preferred that the DNA fragments so produced are in the range of 200-400 — base pairs, a size sufficient to ensure homologous recombination and to insure a representative library in *E.coli*. The fragments are then inserted into appropriately tagged plasmids as described in Hensel et al., Science 269: 400-403(1995). Although a number of plasmids can be used for this purpose, a particularly useful plasmid is pJDC9 described by Pearce et al., Mol. Microbiol. 9:1037 (1993) which carries the *erm* gene facilitating erythromycin selection in either *E. coli* or *S. pneumoniae* previously modified by incorporation of DNA sequence tags into one of the polylinker cloning sites. The tagged plasmids are introduced into the appropriate *S. pneumoniae* strain selected, *inter alia*, on the basis of serotype and virulence in a murine model of pneumococcal pneumonia.

It is appreciated that a seventeen amino acid competence factor exists (Havastein et al., Proc. Nat'l. Acad. Sci. USA 92:11140-44 (1995)) and may be usefully employed in this protocol to increase the transformation frequencies. A proportion of transformants are analysed to verify homologous integration and as a check on stability. Unwanted levels of reversion are minimized because the duplicated regions will be short (200-400 bp), however if significant reversion rates are encountered they may be modulated by maintaining antibiotic selection during the growth of the transformants in culture and/or during growth in the animal.

The *S. pneumoniae* transformants are pooled for inoculation into mice, *eg.*, Swiss and/or C57B1/6. Preliminary experiments are conducted to establish the optimum complexity of the pools and level of inoculum. A particularly useful model has been described by Veber et al. (J. Antimicrobiol. Chemother. 32:432 (1993) in which 10^5 cfu inocula sizes are introduced by mouth to the trachea. Strain differences are observed with respect to onset of disease *e.g.*, 3-4 days for Swiss mice and 8-10 days for C57B1/6. Infection yields in the lungs approach 10^8 cfu/lung. IP administration is also possible when genes mediating blood stream infection are evaluated. Following optimization of parameters of the infection model, the mutant bank normally comprising several thousand strains is subjected to the virulence test. Mutants with attenuated virulence are identified by hybridization analysis using the labelled tags from the "input" and "recovered" pools as probes as described in Hensel et al., Science 269: 400-403(1995). *S. pneumoniae* DNA is colony blotted or dot blotted, DNA flanking the integrated plasmid is cloned by plasmid rescue in *E. coli* (Morrison et al., J. Bacteriol. 159:870 (1984)) and sequenced. Following sequencing, the DNA is compared to the nucleotide sequences given herein and the appropriate ORF is identified and function confirmed for example by knock-out studies.

Expression vectors providing the selected protein are prepared and the protein is configured — in an appropriate screen for the identification of anti-microbial agents. Alternatively, genomic DNA libraries are probed with restriction fragments flanking the integrated plasmid to isolate full-length cloned virulence genes whose function can be confirmed by "knock-out" studies or other methods, which are then expressed and incorporated into a screen as described above.

What is claimed is 1. An isolated polynucleotide comprising a polynucleotide — sequence selected from the group consisting of:

(a) a polynucleotide having at least a 70% identity to a polynucleotide encoding a polypeptide comprising an amino acid sequence of Table 1;

(b) a polynucleotide having at least a 70% identity to a polynucleotide encoding a mature polypeptide expressed by the gene contained in the *S. pneumoniae* of the deposited strain that was sequenced to obtain a polynucleotide sequence of Table 1;

(c) a polynucleotide encoding a polypeptide comprising an amino acid sequence which is at least 70% identical to an amino acid sequence of Table 1;

(d) a polynucleotide which is complementary to the polynucleotide of (a), (b) or (c); and

(e) a polynucleotide comprising at least 15 sequential bases of the polynucleotide of (a), (b), (c) or (d).

2. The polynucleotide of Claim 1 wherein the polynucleotide is DNA.

3. The polynucleotide of Claim 1 wherein the polynucleotide is RNA.

4. The polynucleotide of Claim 2 comprising the nucleic acid sequence selected from the group consisting of the nucleic acid sequences set forth in Table 1.

5. The polynucleotide of Claim 2 which encodes a polypeptide comprising an amino acid sequence selected from the group consisting of the amino acid sequences set forth in Table 1.

6. A vector comprising the polynucleotide of Claim 1.

7. A host cell comprising the vector of Claim 6.

8. A process for producing a polypeptide comprising: expressing from the host cell of Claim 7 a polypeptide encoded by said DNA.

9. A process for producing a polypeptide or fragment comprising culturing a host of claim 7 under conditions sufficient for the production of said polypeptide or fragment.

10. A polypeptide comprising an amino acid sequence which is at least 70% identical to an amino acid sequence selected from the group consisting of the amino acid sequences set forth in Table 1.

11. A polypeptide comprising an amino acid sequence selected from the group consisting of the amino acid sequences set forth in Table 1.

12. An antibody against the polypeptide of claim 10.

13. An antagonist or agonist of the activity or expression of the polypeptide of claim 10.

14. A method for the treatment or prevention of disease of an individual comprising: administering to the individual a therapeutically effective amount of the polypeptide of claim 10.

15. A method for the treatment of an individual having need to inhibit a bacterial polypeptide comprising: administering to the individual a therapeutically effective amount of the antagonist of Claim 13.

16. A process for diagnosing a disease related to expression or activity of the polypeptide of claim 10 in an individual comprising:

- (a) determining a nucleic acid sequence encoding said polypeptide, and/or
- (b) analyzing for the presence or amount of said polypeptide in a sample derived from the individual.

17. A method for identifying compounds which interact with and inhibit or activate an activity of the polypeptide of claim 10 comprising:

contacting a composition comprising the polypeptide with the compound to be screened under conditions to permit interaction between the compound and the polypeptide to assess the interaction of a compound, such interaction being associated with a second component capable of providing a detectable signal in response to the interaction of the polypeptide with the compound;

and determining whether the compound interacts with and activates or inhibits an activity of the polypeptide by detecting the presence or absence of a signal generated from the interaction of the compound with the polypeptide.

18. A method for inducing an immunological response in a mammal which comprises inoculating the mammal with the polypeptide of claim 10, or a fragment or variant thereof, adequate to produce antibody and/or T cell immune response to protect said animal from disease.

19. A method of inducing immunological response in a mammal which comprises delivering a nucleic acid vector to direct expression of a polypeptide of claim 10, or fragment or a variant thereof, for expressing said polypeptide, or a fragment or a variant thereof *in vivo* in order to induce an immunological response to produce antibody and/ or T cell immune response to protect said animal from disease.

20. A polynucleotide comprising a polynucleotide sequence selected from the group consisting of the the first ten polynucleotides sequences from the top of Table 1.

21. A polypeptide comprising a polypeptide encoded by the polynucleotide of — claim 20.

22. The isolated polynucleotide of claim 1 wherein said nucleotide is selected from the group consisting of:

(a) a polynucleotide having at least a 90% identity to a polynucleotide encoding a polypeptide comprising the amino acid sequence of Table 1;

(b) a polynucleotide having at least a 90% identity to a polynucleotide encoding the same mature polypeptide expressed by the gene contained in the *S. pneumoniae* of the deposited strain that was sequenced to obtain a polynucleotide sequence of Table 1;

(c) a polynucleotide encoding a polypeptide comprising an amino acid sequence which is at least 90% identical to the amino acid sequence of Table 1;

(d) a polynucleotide which is complementary to the polynucleotide of (a), (b) or (c); and

(e) a polynucleotide comprising at least 15 sequential bases of the polynucleotide of (a), (b), (c) or (d).

23. The isolated polynucleotide of claim 1 selected from the group consisting of:

(a) a polynucleotide having at least a 95% identity to a polynucleotide encoding a polypeptide comprising the amino acid sequence of Table 1;

(b) a polynucleotide having at least a 95% identity to a polynucleotide encoding the same mature polypeptide expressed by the gene contained in the *S. pneumoniae* of the deposited strain that was sequenced to obtain a polynucleotide sequence of Table 1;

(c) a polynucleotide encoding a polypeptide comprising an amino acid sequence which is at least 95% identical to the amino acid sequence of Table 1;

(d) a polynucleotide which is complementary to the polynucleotide of (a), (b) or (c); and

(e) a polynucleotide comprising at least 15 sequential bases of the polynucleotide of (a), (b), (c) or (d).

24. An isolated polynucleotide comprising a polynucleotide sequence selected from the group consisting of:

(a) a polynucleotide having at least a 50% identity to a polynucleotide encoding a polypeptide comprising the amino acid sequence of Table 1 and obtained from a prokaryotic species other than *S. pneumoniae*;

(b) a polynucleotide encoding a polypeptide comprising an amino acid sequence which is at least 50% identical to the amino acid sequence of Table 1 and obtained from a prokaryotic species other than *S. pneumoniae*; and

(c) a polynucleotide which is complementary to the polynucleotide of (a) or (b).

25. An isolated Streptococcal polypeptide having one of the amino acid sequences given in Table 1.

26. An isolated nucleic acid encoding one of the amino acid sequences of Claim 1 and nucleic acid sequences capable of hybridizing therewith under stringent conditions.

27. Recombinant vectors comprising the nucleic acid sequences of Claim 26 and host cells transformed or transfected therewith.

28. A method of identifying an antimicrobial compound comprising contacting candidate compounds with a polypeptide of Claim 1 and selecting those compounds capable of inhibiting the bioactivity of said polypeptide.

29. Antimicrobial compounds identified by the method of Claim 28.

30. An isolated Streptococcal polypeptide having one of the amino acid sequences given in Table 1.

31. An isolated nucleic acid encoding one of the amino acid sequences of Claim 30 and nucleic acid sequences capable of hybridizing therewith under stringent conditions.

32. Recombinant vectors comprising the nucleic acid sequences of Claim 31 and host cells transformed or transfected therewith.

33. A method of identifying an antimicrobial compound comprising contacting candidate compounds with a polypeptide of Claim 30 and selecting those compounds capable of inhibiting the bioactivity of said polypeptide.

34. Antimicrobial compounds identified by the method of Claim 33.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US97/19226

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) :Please See Extra Sheet.

US CL :Please See Extra Sheet.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 424/130.1, 139.1, 184.1; 435/7.1, 69.3, 320.1, 325; 530/300, 350; 536/23.7

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

APS, MEDLINE, BIOSIS, CA, EMBASE, WPIDS

terms: Streptococcus, pneumoniae, dna, polypeptide, treat, diagnose

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 5,476,929 A (BRILES ET AL) 19 December 1995, see entire document	1-5, 20, 22-24, 26, 31
A	SEVIER et al. Monoclonal Antibodies in Clinical Immunology. Clinical Chemistry. 1981, Vol. 27, No. 11, pages 1797-1806, see entire document	1-34
A	US 4,601,980 A (GOEDDEL ET AL) 22 July 1986, see entire document.	1-34
A	US 5,474,905 A (TAI ET AL) 12 December 1995, see entire document.	1-34

☐ Further documents are listed in the continuation of Box C.
 ☐ See patent family annex.

* Special categories of cited documents:		"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E"	earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search

21 JANUARY 1998

Date of mailing of the international search report

20 FEB 1998

 Name and mailing address of the ISA/US
 Commissioner of Patents and Trademarks
 Box PCT
 Washington, D.C. 20231

Facsimile No. (703) 305-3230

 Authorized officer
 MARK NAVARRO

Telephone No. (703) 308-0196

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US97/19226

A. CLASSIFICATION OF SUBJECT MATTER:
IPC (6):

A61K 38/00, 39/00, 39/395, 39/40; C07H 21/04; C07K 1/00; C12N 15/00; C12P 21/06; G01N 33/53

A. CLASSIFICATION OF SUBJECT MATTER:
US CL :

424/130.1, 139.1, 184.1; 435/7.1, 69.3, 320.1, 325; 530/300, 350; 536/23.7



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<p>(51) International Patent Classification ⁶ : A61K 38/00, 39/00, 39/395, 39/40, C07H 21/04, C07K 1/00, C12N 15/00, C12P 21/06, G01N 33/53</p>	<p>A1</p>	<p>(11) International Publication Number: WO 98/19689</p> <p>(43) International Publication Date: 14 May 1998 (14.05.98)</p>
<p>(21) International Application Number: PCT/US97/19226</p> <p>(22) International Filing Date: 27 October 1997 (27.10.97)</p> <p>(30) Priority Data: 60/029,930 1 November 1996 (01.11.96) US</p> <p>(71) Applicants (for all designated States except US): SMITHK-LINE BEECHAM CORPORATION [US/US]; One Franklin Plaza, Philadelphia, PA 19103 (US). SMITHK-LINE BEECHAM PLC [GB/GB]; New Horizons Court, Brentford, Middlesex TW8 9EP (GB).</p> <p>(72) Inventors; and (75) Inventors/Applicants (for US only): BLACK, Michael, Ter-ance [GB/US]; 502 Milhouse Way, Chester Springs, PA 19425 (US). HODGSON, John, Edward [GB/US]; 260 Lapp Road, Malvern, PA 19355 (US). KNOWLES, David, Justin, Charles [GB/GB]; Downsvie House, 45 Cronks Hill Road, Redhill, Surrey RH1 6LY (GB). LONETTO, Michael, Arthur [US/US]; 18 Victoria Circle, Collegeville, PA 19426 (US). NICHOLAS, Richard, Oakley [GB/US]; 355 Carmen Drive, Collegeville, PA 19426 (US). REID, Robert, H., Jr. [US/US]; 8 Pacer Lane, East Norriton, PA 19401 (US).</p>		<p>ZARFOS, Phillip, N. [US/US]; 1907 Yorktown North, Nor-ristown, PA 19403 (US).</p> <p>(74) Agents: GIMMI, Edward, R. et al.; SmithKline Beecham Corporation, Corporate Intellectual Property, UW2220, 709 Swedeland Road, P.O. Box 1539, King of Prussia, PA 19406-0939 (US).</p> <p>(81) Designated States: CA, JP, US, European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).</p> <p>Published With international search report.</p>
<p>(54) Title: NOVEL CODING SEQUENCES</p>		
<p>(57) Abstract</p> <p>This invention relates to newly identified Streptococcal polynucleotides, polypeptides encoded by such polynucleotides, the uses of such polynucleotides and polypeptides, as well as the production of such polynucleotides and polypeptides and recombinant host cells transformed with the polynucleotides. This invention also relates to inhibiting the biosynthesis or action of such polynucleotides or polypeptides and to the use of such inhibitors in therapy.</p>		

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakistan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

NOVEL CODING SEQUENCES

FIELD OF THE INVENTION

This invention relates to newly identified polynucleotides and polypeptides, and their production and uses, as well as their variants, agonists and antagonists, and their uses. In particular, in these and in other regards, the invention relates to novel polynucleotides and polypeptides set forth in Table 1.

BACKGROUND OF THE INVENTION

The Streptococci make up a medically important genera of microbes known to cause several types of disease in humans, including otitis media, pneumonia and meningitis. Since its isolation more than 100 years ago, *Streptococcus pneumoniae* (herein *S. pneumoniae*) has been one of the more intensively studied microbes. For example, much of our early understanding that DNA is, in fact, the genetic material was predicated on the work of Griffith and of Avery, Macleod and McCarty using this microbe. Despite the vast amount of research with *S. pneumoniae*, many questions concerning the virulence of this microbe remain.

While certain Streptococcal factors associated with pathogenicity have been identified, e.g., capsule polysaccharides, peptidoglycans, pneumolysins, PspA Complement factor H binding component, autolysin, neuraminidase, peptide permeases, hydrogen peroxide, IgA1 protease, the list is certainly not complete. Further very little is known concerning the temporal expression of such genes during infection and disease progression in a mammalian host. Discovering the sets of genes the bacterium is likely to be expressing at the different stages of infection, particularly when an infection is established, provides critical information for the screening and characterization of novel antibacterials which can interrupt pathogenesis. In addition to providing a fuller understanding of known proteins, such an approach will identify previously unrecognised targets.

GUG is used as an initiating nucleotide, rather than ATG, for a significant number of mRNA's in both Gram positive and Gram negative bacteria. Statistics on the frequency of NTG codons in the start codon for several bacterial species are available on line via computer at http://biochem.otago.ac.nz:800/Transterm/home_page.html).

A discussion of initiation codons in *B. subtilis* is set forth in Vellanoweth, RL.1993 in *Bacillus subtilis* and other Gram Positive Bacteria, Biochemistry, Physiology and Molecular Genetics, Sonenshein, Hoch, Losick Eds. Amer. Soc. Microbiol, Washington DC. p. 699-711. Vellenworth indicates a major difference between *B. subtilis* and the

gram-negative organisms is in the choice of initiation codon. 91% of the sequenced *E. coli* genes start with AUG. By contrast, about 30% of *B. subtilis* and other clostridial branch genes start with UUG or GUG. Moreover, CUG functions as a start codon in *B. subtilis*. Mutations of an AUG initiation codon to GUG or UUG often cause decreased expression in *B. subtilis* and *E. coli*. Generally, translation efficiency is higher with AUG initiation codons. A strong Shine-Delgarno ribosome binding site, however, can compensate almost fully for a weak initiation codon. It has been reported that genes with a range of expression levels have initiation codons other than ATG in gram positives (Vellanoweth, RL.1993 in Bacillus subtilis and other Gram Positive Bacteria, Biochemistry, Physiology and Molecular Genetics, Sonenshein, Hoch, Losick Eds. Amer. Soc. Microbiol, Washington DC. p. 699-711).

Provided herein are ORF sequences from genes possessing GUG initiation codons and proteins expressed therefrom and homologues thereto to be used for screening for antimicrobial compounds. Clearly, there is a need for polypeptide and polynucleotide sequences that may be used to screen for antimicrobial compound and which may also be used to determine the roles of such sequences in pathogenesis of infection, dysfunction and disease. There is also need, therefore, for identification and characterization of such sequences which may play a role in preventing, ameliorating or correcting infections, dysfunctions or diseases.

The polypeptides of the invention have amino acid sequence homology to a known protein(s) as set forth in Table 1.

SUMMARY OF THE INVENTION

It is an object of the invention to provide polypeptides that have been identified as novel polypeptides by homology between an amino acid sequence selected from the group consisting of the sequences set out in Table 1 and a known amino acid sequence or sequences of other proteins such as the protein identities listed in Table 1.

It is a further object of the invention to provide polynucleotides that encode novel polypeptides, particularly polynucleotides that encode polypeptides of *Streptococcus pneumoniae*.

In a particularly preferred embodiment of the invention the polynucleotide comprises a region encoding a polypeptide comprising a sequence selected from the group consisting of the sequences set out in Table 1, or a variant of any of these sequences.

In another particularly preferred embodiment of the invention there is a novel protein from *Streptococcus pneumoniae* comprising an amino acid sequence selected from the group consisting of the sequences set out in Table 1, or a variant of any of these sequences.

In accordance with another aspect of the invention there is provided an isolated nucleic acid molecule encoding a mature polypeptide expressible by the *Streptococcus pneumoniae* 0100993 strain contained in the deposited strain.

A further aspect of the invention there are provided isolated nucleic acid molecules encoding a polypeptide of the invention, particularly *Streptococcus pneumoniae* polypeptide, and including mRNAs, cDNAs, genomic DNAs. Further embodiments of the invention include biologically, diagnostically, prophylactically, clinically or therapeutically useful variants thereof, and compositions comprising the same.

In accordance with another aspect of the invention, there is provided the use of a polynucleotide of the invention for therapeutic or prophylactic purposes, in particular genetic immunization. Among the particularly preferred embodiments of the invention are naturally occurring allelic variants of a polypeptide of the invention and polypeptides encoded thereby.

Another aspect of the invention there are provided novel polypeptides of *Streptococcus pneumoniae* as well as biologically, diagnostically, prophylactically, clinically or therapeutically useful variants thereof, and compositions comprising the same.

Among the particularly preferred embodiments of the invention are variants of the polypeptides of the invention encoded by naturally occurring alleles of their genes.

In a preferred embodiment of the invention there are provided methods for producing the aforementioned polypeptides.

In accordance with yet another aspect of the invention, there are provided inhibitors to such polypeptides, useful as antibacterial agents, including, for example, antibodies.

In accordance with certain preferred embodiments of the invention, there are provided products, compositions and methods for assessing expression of the polypeptides and polynucleotides of the invention, treating disease, for example, including, for example, otitis media, conjunctivitis, pneumonia, bacteremia, meningitis, sinusitis, pleural empyema and endocarditis, and most particularly meningitis, such as for example infection of cerebrospinal fluid, assaying genetic variation, and administering a polypeptide or polynucleotide of the

invention to an organism to raise an immunological response against a bacteria, especially a *Streptococcus pneumoniae* bacteria.

In accordance with certain preferred embodiments of this and other aspects of the invention there are provided polynucleotides that hybridize to a polynucleotide sequence of the invention, particularly under stringent conditions.

In certain preferred embodiments of the invention there are provided antibodies against polypeptides of the invention.

In other embodiments of the invention there are provided methods for identifying compounds which bind to or otherwise interact with and inhibit or activate an activity of a polypeptide or polynucleotide of the invention comprising: contacting a polypeptide or polynucleotide of the invention with a compound to be screened under conditions to permit binding to or other interaction between the compound and the polypeptide or polynucleotide to assess the binding to or other interaction with the compound, such binding or interaction being associated with a second component capable of providing a detectable signal in response to the binding or interaction of the polypeptide or polynucleotide with the compound; and determining whether the compound binds to or otherwise interacts with and activates or inhibits an activity of the polypeptide or polynucleotide by detecting the presence or absence of a signal generated from the binding or interaction of the compound with the polypeptide or polynucleotide.

In accordance with yet another aspect of the invention, there are provided agonists and antagonists of the polypeptides and polynucleotides of the invention, preferably bacteriostatic or bacteriocidal agonists and antagonists.

In a further aspect of the invention there are provided compositions comprising a polynucleotide or a polypeptide of the invention for administration to a cell or to a multicellular organism.

Various changes and modifications within the spirit and scope of the disclosed invention will become readily apparent to those skilled in the art from reading the following descriptions and from reading the other parts of the present disclosure.

GLOSSARY

The following definitions are provided to facilitate understanding of certain terms used frequently herein.

"Disease(s)" means any bacterial infection, but preferably a streptococcal infection, such as, otitis media, conjunctivitis, pneumonia, bacteremia, meningitis, sinusitis, pleural empyema, endocarditis, meningitis, and infection of cerebrospinal fluid.

"Host cell" is a cell which has been transformed or transfected, or is capable of transformation or transfection by an exogenous polynucleotide sequence.

"Identity," as known in the art, is a relationship between two or more polypeptide sequences or two or more polynucleotide sequences, as determined by comparing the sequences. In the art, "identity" also means the degree of sequence relatedness between polypeptide or polynucleotide sequences, as the case may be, as determined by the match between strings of such sequences. "Identity" and "similarity" can be readily calculated by known methods, including but not limited to those described in (*Computational Molecular Biology*, Lesk, A.M., ed., Oxford University Press, New York, 1988; *Biocomputing: Informatics and Genome Projects*, Smith, D.W., ed., Academic Press, New York, 1993; *Computer Analysis of Sequence Data*, Part I, Griffin, A.M., and Griffin, H.G., eds., Humana Press, New Jersey, 1994; *Sequence Analysis in Molecular Biology*, von Heinje, G., Academic Press, 1987; and *Sequence Analysis Primer*, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, 1991; and Carillo, H., and Lipman, D., *SIAM J. Applied Math.*, 48: 1073 (1988). Preferred methods to determine identity are designed to give the largest match between the sequences tested. Methods to determine identity and similarity are codified in publicly available computer programs. Preferred computer program methods to determine identity and similarity between two sequences include, but are not limited to, the GCG program package (Devereux, J., et al., *Nucleic Acids Research* 12(1): 387 (1984)), BLASTP, BLASTN, and FASTA (Atschul, S.F. et al., *J. Molec. Biol.* 215: 403-410 (1990)). The BLAST X program is publicly available from NCBI and other sources (*BLAST Manual*, Altschul, S., et al., NCBI NLM NIH Bethesda, MD 20894; Altschul, S., et al., *J. Mol. Biol.* 215: 403-410 (1990)). As an illustration, by a polynucleotide having a nucleotide sequence having at least, for example, 95% "identity" to a reference nucleotide sequence it is intended that the nucleotide sequence of the tested polynucleotide is identical to the reference sequence except that the polynucleotide sequence may include up to five point mutations per each 100 nucleotides of the reference nucleotide sequence. In other words, to obtain a polynucleotide having a nucleotide sequence at least 95% identical to a reference nucleotide sequence, up to 5% of the nucleotides in the reference sequence may be deleted or substituted with another

nucleotide, or a number of nucleotides up to 5% of the total nucleotides in the reference sequence may be inserted into the reference sequence. These mutations of the reference sequence may occur at the 5' or 3' terminal positions of the reference nucleotide sequence or anywhere between those terminal positions, interspersed either individually among nucleotides in the reference sequence or in one or more contiguous groups within the reference sequence. Analogously, by a polypeptide having an amino acid sequence having at least, for example, 95% identity to a reference amino acid sequence is intended that the test amino acid sequence of the polypeptide is identical to the reference sequence except that the polypeptide sequence may include up to five amino acid alterations per each 100 amino acids of the reference amino acid. In other words, to obtain a polypeptide having an amino acid sequence at least 95% identical to a reference amino acid sequence, up to 5% of the amino acid residues in the reference sequence may be deleted or substituted with another amino acid, or a number of amino acids up to 5% of the total amino acid residues in the reference sequence may be inserted into the reference sequence. These alterations of the reference sequence may occur at the amino or carboxy terminal positions of the reference amino acid sequence or anywhere between those terminal positions, interspersed either individually among residues in the reference sequence or in one or more contiguous groups within the reference sequence.

"Isolated" means altered "by the hand of man" from its natural state, *i.e.*, if it occurs in nature, it has been changed or removed from its original environment, or both. For example, a polynucleotide or a polypeptide naturally present in a living organism is not "isolated," but the same polynucleotide or polypeptide separated from the coexisting materials of its natural state is "isolated", as the term is employed herein.

"Polynucleotide(s)" generally refers to any polyribonucleotide or polydeoxribonucleotide, which may be unmodified RNA or DNA or modified RNA or DNA. "Polynucleotide(s)" include, without limitation, single- and double-stranded DNA, DNA that is a mixture of single- and double-stranded regions or single-, double- and triple-stranded regions, single- and double-stranded RNA, and RNA that is mixture of single- and double-stranded regions, hybrid molecules comprising DNA and RNA that may be single-stranded or, more typically, double-stranded, or triple-stranded regions, or a mixture of single- and double-stranded regions. In addition, "polynucleotide" as used herein refers to triple-stranded regions comprising RNA or DNA or both RNA and DNA. The strands in such regions may be from the same molecule or from different molecules. The regions may include all of one or more

of the molecules, but more typically involve only a region of some of the molecules. One of the molecules of a triple-helical region often is an oligonucleotide. As used herein, the term "polynucleotide(s)" also includes DNAs or RNAs as described above that contain one or more modified bases. Thus, DNAs or RNAs with backbones modified for stability or for other reasons are "polynucleotide(s)" as that term is intended herein. Moreover, DNAs or RNAs comprising unusual bases, such as inosine, or modified bases, such as tritylated bases, to name just two examples, are polynucleotides as the term is used herein. It will be appreciated that a great variety of modifications have been made to DNA and RNA that serve many useful purposes known to those of skill in the art. The term "polynucleotide(s)" as it is employed herein embraces such chemically, enzymatically or metabolically modified forms of polynucleotides, as well as the chemical forms of DNA and RNA characteristic of viruses and cells, including, for example, simple and complex cells. "Polynucleotide(s)" also embraces short polynucleotides often referred to as oligonucleotide(s).

"Polypeptide(s)" refers to any peptide or protein comprising two or more amino acids joined to each other by peptide bonds or modified peptide bonds. "Polypeptide(s)" refers to both short chains, commonly referred to as peptides, oligopeptides and oligomers and to longer chains generally referred to as proteins. Polypeptides may contain amino acids other than the 20 gene encoded amino acids. "Polypeptide(s)" include those modified either by natural processes, such as processing and other post-translational modifications, but also by chemical modification techniques. Such modifications are well described in basic texts and in more detailed monographs, as well as in a voluminous research literature, and they are well known to those of skill in the art. It will be appreciated that the same type of modification may be present in the same or varying degree at several sites in a given polypeptide. Also, a given polypeptide may contain many types of modifications. Modifications can occur anywhere in a polypeptide, including the peptide backbone, the amino acid side-chains, and the amino or carboxyl termini. Modifications include, for example, acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphatidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent cross-links, formation of cysteine, formation of pyroglutamate, formylation, gamma-carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, proteolytic processing, phosphorylation, prenylation, racemization, glycosylation,

lipid attachment, sulfation, gamma-carboxylation of glutamic acid residues, hydroxylation and ADP-ribosylation, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins, such as arginylation, and ubiquitination. See, for instance, *PROTEINS - STRUCTURE AND MOLECULAR PROPERTIES*, 2nd Ed., T. E. Creighton, W. H. Freeman and Company, New York (1993) and Wold, F., *Posttranslational Protein Modifications: Perspectives and Prospects*, pgs. 1-12 in *POSTTRANSLATIONAL COVALENT MODIFICATION OF PROTEINS*, B. C. Johnson, Ed., Academic Press, New York (1983); Seifter et al., *Meth. Enzymol.* 182:626-646 (1990) and Rattan et al., *Protein Synthesis: Posttranslational Modifications and Aging*, Ann. N.Y. Acad. Sci. 663: 48-62 (1992). Polypeptides may be branched or cyclic, with or without branching. Cyclic, branched and branched circular polypeptides may result from post-translational natural processes and may be made by entirely synthetic methods, as well.

"Variant(s)" as the term is used herein, is a polynucleotide or polypeptide that differs from a reference polynucleotide or polypeptide respectively, but retains essential properties. A typical variant of a polynucleotide differs in nucleotide sequence from another, reference polynucleotide. Changes in the nucleotide sequence of the variant may or may not alter the amino acid sequence of a polypeptide encoded by the reference polynucleotide. Nucleotide changes may result in amino acid substitutions, additions, deletions, fusions and truncations in the polypeptide encoded by the reference sequence, as discussed below. A typical variant of a polypeptide differs in amino acid sequence from another, reference polypeptide. Generally, differences are limited so that the sequences of the reference polypeptide and the variant are closely similar overall and, in many regions, identical. A variant and reference polypeptide may differ in amino acid sequence by one or more substitutions, additions, deletions in any combination. A substituted or inserted amino acid residue may or may not be one encoded by the genetic code. A variant of a polynucleotide or polypeptide may be a naturally occurring such as an allelic variant, or it may be a variant that is not known to occur naturally. Non-naturally occurring variants of polynucleotides and polypeptides may be made by mutagenesis techniques, by direct synthesis, and by other recombinant methods known to skilled artisans.

DESCRIPTION OF THE INVENTION

Each of polynucleotide and polypeptide sequences provided herein may be used in the discovery and development of antibacterial compounds. Upon expression of the sequences with the appropriate initiation and termination codons the encoded polypeptide

can be used as a target for the screening of antimicrobial drugs. Additionally, the DNA sequences encoding preferably the amino terminal regions of the encoded protein or the Shine-Delgarno region can be used to construct antisense sequences to control the expression of the coding sequence of interest. Furthermore, many of the sequences disclosed herein also provide regions upstream and downstream from the encoding sequence. These sequences are useful as a source of regulatory elements for the control of bacterial gene expression. Such sequences are conveniently isolated by restriction enzyme action or synthesized chemically and introduced, for example, into promoter identification strains. These strains contain a reporter structural gene sequence located downstream from a restriction site such that if an active promoter is inserted, the reporter gene will be expressed.

Although each of the sequences may be employed as described above, this invention also provides several means for identifying particularly useful target genes. The first of these approaches entails searching appropriate databases for sequence matches in related organisms. Thus, if a homologue exists, the Streptococcal-like form of this gene would likely play an analogous role. For example, a Streptococcal protein identified as homologous to a cell surface protein in another organism would be useful as a vaccine candidate. To the extent such homologies have been identified for the sequences disclosed herein they are reported along with the encoding sequence.

Each of the DNA sequences provided herein may be used in the discovery and development of antibacterial compounds. Because each of the sequences contains an open reading frame (ORF) with an appropriate initiation and termination codons, the encoded protein upon expression can be used as a target for the screening of antimicrobial drugs. Additionally, the DNA sequences encoding the amino terminal regions of the encoded protein can be used to construct antisense sequences to control the expression of the coding sequence of interest. Furthermore, many of the sequences disclosed herein also provide regions upstream and downstream from the encoding sequence. These sequences are useful as a source of regulatory elements for the control of bacterial gene expression. Such sequences are conveniently isolated by restriction enzyme action or synthesized chemically and introduced, for example, into promoter identification strains. These strains contain a reporter structural gene sequence located downstream from a restriction site such that if an active promoter is inserted, the reporter gene will be expressed.

It is believed that bacteria possess a number of ways of regulating gene expression levels, especially in subtle degrees, and the interplay between ribosome binding site and initiation codon is utilized for this purpose for these genes. It is also believed that such genes will be important targets for antimicrobial drug discovery, particularly since pathogenesis genes are believed undergo gene expression regulation during in the pathogenesis process. Therefore, the invention provides ORF sequences possessing a GTG (GUG) initiation codon and protein targets expressed therefrom.

Although each of the sequences may be employed as described above, this invention also provides several means for identifying particularly useful target genes. The first of these approaches entails searching appropriate databases for sequence matches in related organisms. Thus, if a homologue exists, the Streptococcal-like form of this gene would likely play an analogous role. For example, a Streptococcal protein identified as homologous to a cell surface protein in another organism would be useful as a vaccine candidate. To the extent such homologies have been identified for the sequences disclosed herein they are reported along with the encoding sequence.

ORF Gene Expression

Recently techniques have become available to evaluate temporal gene expression in bacteria, particularly as it applies to viability under laboratory and infection conditions. A number of methods can be used to identify genes which are essential to survival *per se*, or essential to the establishment/maintenance of an infection. Identification of an ORF unknown by one of these methods yields additional information about its function and permits the selection of such an ORF for further development as a screening target. Briefly, these approaches include:

1) **Signature Tagged Mutagenesis (STM)**: This technique is described by Hensel *et al.*, Science 269: 400-403(1995), the contents of which is incorporated by reference for background purposes. Signature tagged mutagenesis identifies genes necessary for the establishment/maintenance of infection in a given infection model.

The basis of the technique is the random mutagenesis of target organism by various means (e.g., transposons) such that unique DNA sequence tags are inserted in close proximity to the site of mutation. The tags from a mixed population of bacterial mutants and bacteria recovered from an infected hosts are detected by amplification, radiolabeling and hybridisation analysis. Mutants attenuated in virulence are revealed by absence of the tag from the pool of bacteria recovered from infected hosts.

In *Streptococcus pneumoniae*, because the transposon system is less well developed, a more efficient way of creating the tagged mutants is to use the insertion-duplication mutagenesis technique as described by Morrison et al., J. Bacteriol. 159:870 (1984) the contents of which is incorporated by reference for background purposes.

2) In Vivo Expression Technology (IVET): This technique is described by Camilli et al., Proc. Nat'l. Acad. Sci. USA. 91:2634-2638 (1994), the contents of which is incorporated by reference for background purposes. IVET identifies genes up-regulated during infection when compared to laboratory cultivation, implying an important role in infection. ORF identified by this technique are implied to have a significant role in infection establishment/maintenance.

In this technique random chromosomal fragments of target organism are cloned upstream of a promoter-less recombinase gene in a plasmid vector. This construct is introduced into the target organism which carries an antibiotic resistance gene flanked by resolvase sites. Growth in the presence of the antibiotic removes from the population those fragments cloned into the plasmid vector capable of supporting transcription of the recombinase gene and therefore have caused loss of antibiotic resistance. The resistant pool is introduced into a host and at various times after infection bacteria may be recovered and assessed for the presence of antibiotic resistance. The chromosomal fragment carried by each antibiotic sensitive bacterium should carry a promoter or portion of a gene normally upregulated during infection. Sequencing upstream of the recombinase gene allows identification of the up regulated gene.

3) Differential display: This technique is described by Chuang et al., J. Bacteriol. 175:2026-2036 (1993), the contents of which is incorporated by reference for background purposes. This method identifies those genes which are expressed in an organism by identifying mRNA present using randomly-primed RT-PCR. By comparing pre-infection and post infection profiles, genes up and down regulated during infection can be identified and the RT-PCR product sequenced and matched to ORF 'unknowns'.

4) Generation of conditional lethal mutants by transposon mutagenesis:

This technique, described by de Lorenzo, V. et al., Gene 123:17-24 (1993); Neuwald, A. F. et al., Gene 125: 69-73(1993); and Takiff, H. E. et al., J. Bacteriol. 174:1544-1553(1992), the contents of which is incorporated by reference for background purposes, identifies genes whose expression are essential for cell viability.

In this technique transposons carrying controllable promoters, which provide transcription outward from the transposon in one or both directions, are generated. Random insertion of these transposons into target organisms and subsequent isolation of insertion mutants in the presence of inducer of promoter activity ensures that insertions which separate promoter from coding region of a gene whose expression is essential for cell viability will be recovered. Subsequent replica plating in the absence of inducer identifies such insertions, since they fail to survive. Sequencing of the flanking regions of the transposon allows identification of site of insertion and identification of the gene disrupted. Close monitoring of the changes in cellular processes/morphology during growth in the absence of inducer yields information on likely function of the gene. Such monitoring could include flow cytometry (cell division, lysis, redox potential, DNA replication), incorporation of radiochemically labeled precursors into DNA, RNA, protein, lipid, peptidoglycan, monitoring reporter enzyme gene fusions which respond to known cellular stresses.

5) Generation of conditional lethal mutants by chemical mutagenesis:

This technique is described by Beckwith, J., Methods in Enzymology 204: 3-18(1991), the contents of which are incorporated herein by reference for background purposes. In this technique random chemical mutagenesis of target organism, growth at temperature other than physiological temperature (permissive temperature) and subsequent replica plating and growth at different temperature (e.g. 42°C to identify ts, 25°C to identify cs) are used to identify those isolates which now fail to grow (conditional mutants). As above close monitoring of the changes upon growth at the non-permissive temperature yields information on the function of the mutated gene. Complementation of conditional lethal mutation by library from target organism and sequencing of complementing gene allows matching with unknown ORF.

6) RT-PCR: *Streptococcus pneumoniae* messenger RNA is isolated from bacterial infected tissue e.g. 48 hour murine lung infections, and the amount of each mRNA species assessed by reverse transcription of the RNA sample primed with random hexanucleotides

followed by PCR with gene specific primer pairs. The determination of the presence and amount of a particular mRNA species by quantification of the resultant PCR product provides information on the bacterial genes which are transcribed in the infected tissue. Analysis of gene transcription can be carried out at different times of infection to gain a detailed knowledge of gene regulation in bacterial pathogenesis allowing for a clearer understanding of which gene products represent targets for screens for novel antibacterials. Because of the gene specific nature of the PCR primers employed it should be understood that the bacterial mRNA preparation need not be free of mammalian RNA. This allows the investigator to carry out a simple and quick RNA preparation from infected tissue to obtain bacterial mRNA species which are very short lived in the bacterium (in the order of 2 minute half-lives). Optimally the bacterial mRNA is prepared from infected murine lung tissue by mechanical disruption in the presence of TRIzol (GIBCO-BRL) for very short periods of time, subsequent processing according to the manufacturers of TRIzol reagent and DNAase treatment to remove contaminating DNA. Preferably the process is optimised by finding those conditions which give a maximum amount of *Streptococcus pneumoniae* 16S ribosomal RNA as detected by probing Northernblots with a suitably labelled sequence specific oligonucleotide probe. Typically a 5' dye labelled primer is used in each PCR primer pair in a PCR reaction which is terminated optimally between 8 and 25 cycles. The PCR products are separated on 6% polyacrylamide gels with detection and quantification using GeneScanner (manufactured by ABI).

Each of these techniques may have advantages or disadvantages depending on the particular application. The skilled artisan would choose the approach that is the most relevant with the particular end use in mind.

Use of the of these technologies when applied to the ORFs of the present invention enables identification of bacterial proteins expressed during infection, inhibitors of which would have utility in anti-bacterial therapy.

The invention relates to novel polypeptides and polynucleotides as described in greater detail below. In particular, the invention relates to polypeptides and polynucleotides of *Streptococcus pneumoniae*, which is related by amino acid sequence homology to known polypeptide as set forth in Table 1. The invention relates especially to compounds having the nucleotide and amino acid sequence selected from the group consisting of the sequences set

out in Table 1, and to the nucleotide sequences of the DNA in the deposited strain and amino acid sequences encoded thereby.

Deposited materials

The deposit has been made under the terms of the Budapest Treaty on the International Recognition of the Deposit of Micro-organisms for Purposes of Patent Procedure. The strain will be irrevocably and without restriction or condition released to the public upon the issuance of a patent. The deposit is provided merely as convenience to those of skill in the art and is not an admission that a deposit is required for enablement, such as that required under 35 U.S.C. §112.

A deposit containing a *Streptococcus pneumoniae* bacterial strain has been deposited with the National Collections of Industrial and Marine Bacteria Ltd. (NCIMB), 23 St. Machar Drive, Aberdeen AB2 1RY, Scotland on 11 April 1996 and assigned NCIMB Deposit No. 40794. The *Streptococcus pneumoniae* bacterial strain deposit is referred to herein as "the deposited bacterial strain" or as "the DNA of the deposited bacterial strain."

The deposited material is a bacterial strain that contains the full length FabH DNA, referred to as "NCIMB 40794" upon deposit.

The sequence of the polynucleotides contained in the deposited material, as well as the amino acid sequence of the polypeptide encoded thereby, are controlling in the event of any conflict with any description of sequences herein.

A license may be required to make, use or sell the deposited materials, and no such license is hereby granted.

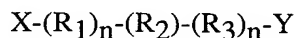
The deposited strain contains the full length genes comprising the polynucleotides set forth in Table 1. The sequence of the polynucleotides contained in the deposited strain, as well as the amino acid sequence of the polypeptide encoded thereby, are controlling in the event of any conflict with any description of sequences herein.

Polypeptides

The polypeptides of the invention include the polypeptides set forth in Table 1 (in particular the mature polypeptide) as well as polypeptides and fragments, particularly those which have the biological activity of a polypeptide of the invention, and also those which have at least 50%, 60% or 70% identity to a polypeptide sequence selected from the group consisting of the sequences set out in Table 1 or the relevant portion, preferably at least 80% identity to a polypeptide sequence selected from the group consisting of the sequences set out in Table 1, and more preferably at least 90% similarity (more preferably at least 90% identity)

to a polypeptide sequence selected from the group consisting of the sequences set out in Table 1, and still more preferably at least 95% similarity (still more preferably at least 95% identity) to a polypeptide sequence selected from the group consisting of the sequences set out in Table 1, and also include portions of such polypeptides with such portion of the polypeptide generally containing at least 30 amino acids and more preferably at least 50 amino acids.

The invention also includes polypeptides of the formula:



wherein, at the amino terminus, X is hydrogen, and at the carboxyl terminus, Y is hydrogen or a metal, R_1 and R_3 are any amino acid residue, n is an integer between 1 and 2000, and R_2 is an amino acid sequence of the invention, particularly an amino acid sequence selected from the group set forth in Table 1. In the formula above R_2 is oriented so that its amino terminal residue is at the left, bound to R_1 , and its carboxy terminal residue is at the right, bound to R_3 . Any stretch of amino acid residues denoted by either R group, where R is greater than 1, may be either a heteropolymer or a homopolymer, preferably a heteropolymer. In preferred embodiments n is an integer between 1 and 1000 or 2000.

A fragment is a variant polypeptide having an amino acid sequence that entirely is the same as part but not all of the amino acid sequence of the aforementioned polypeptides. As with polypeptides, fragments may be "free-standing," or comprised within a larger polypeptide of which they form a part or region, most preferably as a single continuous region, a single larger polypeptide.

Preferred fragments include, for example, truncation polypeptides having a portion of the amino acid sequence of Table 1, or of variants thereof, such as a continuous series of residues that includes the amino terminus, or a continuous series of residues that includes the carboxyl terminus. Degradation forms of the polypeptides of the invention in a host cell, particularly a *Streptococcus pneumoniae*, are also preferred. Further preferred are fragments characterized by structural or functional attributes such as fragments that comprise alpha-helix and alpha-helix forming regions, beta-sheet and beta-sheet-forming regions, turn and turn-forming regions, coil and coil-forming regions, hydrophilic regions, hydrophobic regions, alpha amphipathic regions, beta amphipathic regions, flexible regions, surface-forming regions, substrate binding region, and high antigenic index regions.

Also preferred are biologically active fragments which are those fragments that mediate activities of polypeptides of the invention, including those with a similar activity or an improved activity, or with a decreased undesirable activity. Also included are those

fragments that are antigenic or immunogenic in an animal, especially in a human. Particularly preferred are fragments comprising receptors or domains of enzymes that confer a function essential for viability of *Streptococcus pneumoniae* or the ability to initiate, or maintain cause disease in an individual, particularly a human.

Variants that are fragments of the polypeptides of the invention may be employed for producing the corresponding full-length polypeptide by peptide synthesis; therefore, these variants may be employed as intermediates for producing the full-length polypeptides of the invention.

In addition to the standard single and triple letter representations for amino acids, the term "X" or "Xaa" is also used. "X" and "Xaa" mean that any of the twenty naturally occurring amino acids may appear at such a designated position in the polypeptide sequence.

Polynucleotides

The nucleotide sequences disclosed herein can be obtained by synthetic chemical techniques known in the art or can be obtained from *S. pneumoniae* 0100993 by probing a DNA preparation with probes constructed from the particular sequences disclosed herein. Alternatively, oligonucleotides derived from a disclosed sequence can act as PCR primers in a process of PCR-based cloning of the sequence from a bacterial genomic source. It is recognised that such sequences will also have utility in diagnosis of the stage of infection and type of infection the pathogen has attained.

To obtain the polynucleotide encoding the protein using the DNA sequence given herein typically a library of clones of chromosomal DNA of *S.pneumoniae* 0100993 in *E. coli* or some other suitable host is probed with a radiolabelled oligonucleotide, preferably a 17mer or longer, derived from the partial sequence. Clones carrying DNA identical to that of the probe can then be distinguished using high stringency washes. By sequencing the individual clones thus identified with sequencing primers designed from the original sequence it is then possible to extend the sequence in both directions to determine the full gene sequence. Conveniently such sequencing is performed using denatured double stranded DNA prepared from a plasmid clone. Suitable techniques are described by Maniatis, T., Fritsch, E.F. and Sambrook, J. in MOLECULAR CLONING, A Laboratory Manual, 2nd edition, 1989, Cold Spring Harbor Laboratory (see: Screening By Hybridization 1.90 and Sequencing Denatured Double-Stranded DNA Templates 13.70).

Moreover, another aspect of the invention relates to isolated polynucleotides that encode the polypeptides of the invention having a deduced amino acid sequence selected from

the group consisting of the sequences in Table 1 and polynucleotides closely related thereto and variants thereof.

Using the information provided herein, such as the polynucleotide sequences set out in Table 1, a polynucleotide of the invention encoding polypeptide may be obtained using standard cloning and screening methods, such as those for cloning and sequencing chromosomal DNA fragments from bacteria using *Streptococcus pneumoniae* 0100993 cells as starting material, followed by obtaining a full length clone. For example, to obtain a polynucleotide sequence of the invention, such as a sequence set forth in Table 1, typically a library of clones of chromosomal DNA of *Streptococcus pneumoniae* 0100993 in *E.coli* or some other suitable host is probed with a radiolabeled oligonucleotide, preferably a 17-mer or longer, derived from a partial sequence. Clones carrying DNA identical to that of the probe can then be distinguished using stringent conditions. By sequencing the individual clones thus identified with sequencing primers designed from the original sequence it is then possible to extend the sequence in both directions to determine the full gene sequence. Conveniently, such sequencing is performed using denatured double stranded DNA prepared from a plasmid clone. Suitable techniques are described by Maniatis, T., Fritsch, E.F. and Sambrook et al., *MOLECULAR CLONING, A LABORATORY MANUAL*, 2nd Ed.; Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York (1989). (see in particular Screening By Hybridization 1.90 and Sequencing Denatured Double-Stranded DNA Templates 13.70). Illustrative of the invention, the polynucleotides set out in Table 1 were discovered in a DNA library derived from *Streptococcus pneumoniae* 0100993.

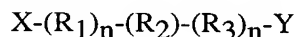
The DNA sequences set out in Table 1 each contains at least one open reading frame encoding a protein having at least about the number of amino acid residues set forth in Table 1. The start and stop codons of each open reading frame (herein "ORF") DNA are the first three and the last three nucleotides of each polynucleotide set forth in Table 1.

Certain polynucleotides and polypeptides of the invention are structurally related to known proteins as set forth in Table 1. These proteins exhibit greatest homology to the homologue listed in Table 1 from among the known proteins.

The invention provides a polynucleotide sequence identical over its entire length to each coding sequence in Table 1. Also provided by the invention is the coding sequence for the mature polypeptide or a fragment thereof, by itself as well as the coding sequence for the mature polypeptide or a fragment in reading frame with other coding sequence, such as those

encoding a leader or secretory sequence, a pre-, or pro- or prepro- protein sequence. The polynucleotide may also contain non-coding sequences, including for example, but not limited to non-coding 5' and 3' sequences, such as the transcribed, non-translated sequences, termination signals, ribosome binding sites, sequences that stabilize mRNA, introns, polyadenylation signals, and additional coding sequence which encode additional amino acids. For example, a marker sequence that facilitates purification of the fused polypeptide can be encoded. In certain embodiments of the invention, the marker sequence is a hexahistidine peptide, as provided in the pQE vector (Qiagen, Inc.) and described in Gentz *et al.*, *Proc. Natl. Acad. Sci., USA* 86: 821-824 (1989), or an HA tag (Wilson *et al.*, *Cell* 37: 767 (1984). Polynucleotides of the invention also include, but are not limited to, polynucleotides comprising a structural gene and its naturally associated sequences that control gene expression.

The invention also includes polynucleotides of the formula:



wherein, at the 5' end of the molecule, X is hydrogen, and at the 3' end of the molecule, Y is hydrogen or a metal, R_1 and R_3 is any nucleic acid residue, n is an integer between 1 and 3000, and R_2 is a nucleic acid sequence of the invention, particularly a nucleic acid sequence selected from the group set forth in Table 1. In the polynucleotide formula above R_2 is oriented so that its 5' end residue is at the left, bound to R_1 , and its 3' end residue is at the right, bound to R_3 . Any stretch of nucleic acid residues denoted by either R group, where R is greater than 1, may be either a heteropolymer or a homopolymer, preferably a heteropolymer. In a preferred embodiment n is an integer between 1 and 1000, or 2000 or 3000.

The term "polynucleotide encoding a polypeptide" as used herein encompasses polynucleotides that include a sequence encoding a polypeptide of the invention, particularly a bacterial polypeptide and more particularly a polypeptide of the *Streptococcus pneumoniae* having an amino acid sequence set out in Table 1. The term also encompasses polynucleotides that include a single continuous region or discontinuous regions encoding the polypeptide (for example, interrupted by integrated phage or an insertion sequence or editing) together with additional regions, that also may contain coding and/or non-coding sequences.

The invention further relates to variants of the polynucleotides described herein that encode for variants of the polypeptide having the deduced amino acid sequence of Table 1. Variants that are fragments of the polynucleotides of the invention may be used to synthesize full-length polynucleotides of the invention.

Further particularly preferred embodiments are polynucleotides encoding polypeptide variants, that have the amino acid sequence of a polypeptide of Table 1 in which several, a few, 5 to 10, 1 to 5, 1 to 3, 2, 1 or no amino acid residues are substituted, deleted or added, in any combination. Especially preferred among these are silent substitutions, additions and deletions, that do not alter the properties and activities of such polynucleotide.

Further preferred embodiments of the invention are polynucleotides that are at least 50%, 60% or 70% identical over their entire length to a polynucleotide encoding a polypeptide having the amino acid sequence set out in Table 1, and polynucleotides that are complementary to such polynucleotides. Alternatively, most highly preferred are polynucleotides that comprise a region that is at least 80% identical over its entire length to a polynucleotide encoding a polypeptide of the deposited strain and polynucleotides complementary thereto. In this regard, polynucleotides at least 90% identical over their entire length to the same are particularly preferred, and among these particularly preferred polynucleotides, those with at least 95% are especially preferred. Furthermore, those with at least 97% are highly preferred among those with at least 95%, and among these those with at least 98% and at least 99% are particularly highly preferred, with at least 99% being the more preferred.

A preferred embodiment is an isolated polynucleotide comprising a polynucleotide sequence selected from the group consisting of: a polynucleotide having at least a 50% identity to a polynucleotide encoding a polypeptide comprising the amino acid sequence of Table 1 and obtained from a prokaryotic species other than *S. pneumoniae*; and a polynucleotide encoding a polypeptide comprising an amino acid sequence which is at least 50% identical to the amino acid sequence of Table 1 and obtained from a prokaryotic species other than *S. pneumoniae*.

Preferred embodiments are polynucleotides that encode polypeptides that retain substantially the same biological function or activity as the mature polypeptide encoded by the DNA of Table 1.

The invention further relates to polynucleotides that hybridize to the herein above-described sequences. In this regard, the invention especially relates to polynucleotides that hybridize under stringent conditions to the herein above-described polynucleotides. As herein used, the terms "stringent conditions" and "stringent hybridization conditions" mean hybridization will occur only if there is at least 95% and preferably at least 97% identity between the sequences. An example of stringent hybridization conditions is overnight

incubation at 42°C in a solution comprising: 50% formamide, 5x SSC (150mM NaCl, 15mM trisodium citrate), 50 mM sodium phosphate (pH7.6), 5x Denhardt's solution, 10% dextran sulfate, and 20 micrograms/ml denatured, sheared salmon sperm DNA, followed by washing the hybridization support in 0.1x SSC at about 65°C. Hybridization and wash conditions are well known and exemplified in Sambrook, *et al.*, Molecular Cloning: A Laboratory Manual, Second Edition, Cold Spring Harbor, N.Y., (1989), particularly Chapter 11 therein.

The invention also provides a polynucleotide consisting essentially of a polynucleotide sequence obtainable by screening an appropriate library containing the complete gene for a polynucleotide sequence set forth in Table 1 under stringent hybridization conditions with a probe having the sequence of said polynucleotide sequence or a fragment thereof; and isolating said DNA sequence. Fragments useful for obtaining such a polynucleotide include, for example, probes and primers described elsewhere herein.

As discussed additionally herein regarding polynucleotide assays of the invention, for instance, polynucleotides of the invention as discussed above, may be used as a hybridization probe for RNA, cDNA and genomic DNA to isolate full-length cDNAs and genomic clones encoding a polypeptide and to isolate cDNA and genomic clones of other genes that have a high sequence similarity to a polynucleotide set forth in Table 1. Such probes generally will comprise at least 15 bases. Preferably, such probes will have at least 30 bases and may have at least 50 bases. Particularly preferred probes will have at least 30 bases and will have 50 bases or less.

For example, the coding region of each gene that comprises or is comprised by a polynucleotide set forth in Table 1 may be isolated by screening using a DNA sequence provided in Table 1 to synthesize an oligonucleotide probe. A labeled oligonucleotide having a sequence complementary to that of a gene of the invention is then used to screen a library of cDNA, genomic DNA or mRNA to determine which members of the library the probe hybridizes to.

The polynucleotides and polypeptides of the invention may be employed, for example, as research reagents and materials for discovery of treatments of and diagnostics for disease, particularly human disease, as further discussed herein relating to polynucleotide assays.

Polynucleotides of the invention that are oligonucleotides derived from the a polynucleotide or polypeptide sequence set forth in Table 1 may be used in the processes

herein as described, but preferably for PCR, to determine whether or not the polynucleotides identified herein in whole or in part are transcribed in bacteria in infected tissue. It is recognized that such sequences will also have utility in diagnosis of the stage of infection and type of infection the pathogen has attained.

The invention also provides polynucleotides that may encode a polypeptide that is the mature protein plus additional amino or carboxyl-terminal amino acids, or amino acids interior to the mature polypeptide (when the mature form has more than one polypeptide chain, for instance). Such sequences may play a role in processing of a protein from precursor to a mature form, may allow protein transport, may lengthen or shorten protein half-life or may facilitate manipulation of a protein for assay or production, among other things. As generally is the case *in vivo*, the additional amino acids may be processed away from the mature protein by cellular enzymes.

A precursor protein, having the mature form of the polypeptide fused to one or more prosequences may be an inactive form of the polypeptide. When prosequences are removed such inactive precursors generally are activated. Some or all of the prosequences may be removed before activation. Generally, such precursors are called proproteins.

In addition to the standard A, G, C, T/U representations for nucleic acid bases, the term "N" is also used. "N" means that any of the four DNA or RNA bases may appear at such a designated position in the DNA or RNA sequence, except it is preferred that N is not a base that when taken in combination with adjacent nucleotide positions, when read in the correct reading frame, would have the effect of generating a premature termination codon in such reading frame.

In sum, a polynucleotide of the invention may encode a mature protein, a mature protein plus a leader sequence (which may be referred to as a preprotein), a precursor of a mature protein having one or more prosequences that are not the leader sequences of a preprotein, or a preproprotein, which is a precursor to a proprotein, having a leader sequence and one or more prosequences, which generally are removed during processing steps that produce active and mature forms of the polypeptide.

Vectors, host cells, expression

The invention also relates to vectors that comprise a polynucleotide or polynucleotides of the invention, host cells that are genetically engineered with vectors of the invention and the production of polypeptides of the invention by recombinant techniques.

Cell-free translation systems can also be employed to produce such proteins using RNAs derived from the DNA constructs of the invention.

For recombinant production, host cells can be genetically engineered to incorporate expression systems or portions thereof or polynucleotides of the invention. Introduction of a polynucleotide into the host cell can be effected by methods described in many standard laboratory manuals, such as Davis et al., *BASIC METHODS IN MOLECULAR BIOLOGY*, (1986) and Sambrook et al., *MOLECULAR CLONING: A LABORATORY MANUAL*, 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1989), such as, calcium phosphate transfection, DEAE-dextran mediated transfection, transvection, microinjection, cationic lipid-mediated transfection, electroporation, transduction, scrape loading, ballistic introduction and infection.

Representative examples of appropriate hosts include bacterial cells, such as streptococci, staphylococci, enterococci *E. coli*, streptomyces and *Bacillus subtilis* cells; fungal cells, such as yeast cells and *Aspergillus* cells; insect cells such as *Drosophila* S2 and *Spodoptera* Sf9 cells; animal cells such as CHO, COS, HeLa, C127, 3T3, BHK, 293 and Bowes melanoma cells; and plant cells.

A great variety of expression systems can be used to produce the polypeptides of the invention. Such vectors include, among others, chromosomal, episomal and virus-derived vectors, *e.g.*, vectors derived from bacterial plasmids, from bacteriophage, from transposons, from yeast episomes, from insertion elements, from yeast chromosomal elements, from viruses such as baculoviruses, papova viruses, such as SV40, vaccinia viruses, adenoviruses, fowl pox viruses, pseudorabies viruses and retroviruses, and vectors derived from combinations thereof, such as those derived from plasmid and bacteriophage genetic elements, such as cosmids and phagemids. The expression system constructs may contain control regions that regulate as well as engender expression. Generally, any system or vector suitable to maintain, propagate or express polynucleotides and/or to express a polypeptide in a host may be used for expression in this regard. The appropriate DNA sequence may be inserted into the expression system by any of a variety of well-known and routine techniques, such as, for example, those set forth in Sambrook *et al.*, *MOLECULAR CLONING, A LABORATORY MANUAL*, (*supra*).

For secretion of the translated protein into the lumen of the endoplasmic reticulum, into the periplasmic space or into the extracellular environment, appropriate secretion signals

may be incorporated into the expressed polypeptide. These signals may be endogenous to the polypeptide or they may be heterologous signals.

Polypeptides of the invention can be recovered and purified from recombinant cell cultures by well-known methods including ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography, and lectin chromatography. Most preferably, high performance liquid chromatography is employed for purification. Well known techniques for refolding protein may be employed to regenerate active conformation when the polypeptide is denatured during isolation and or purification.

Diagnostic Assays

This invention is also related to the use of the polynucleotides of the invention for use as diagnostic reagents. Detection of such polynucleotides in a eukaryote, particularly a mammal, and especially a human, will provide a diagnostic method for diagnosis of a disease. Eukaryotes (herein also "individual(s)"), particularly mammals, and especially humans, infected with an organism comprising a gene of the invention may be detected at the nucleic acid level by a variety of techniques.

Nucleic acids for diagnosis may be obtained from an infected individual's cells and tissues, such as bone, blood, muscle, cartilage, and skin. Genomic DNA may be used directly for detection or may be amplified enzymatically by using PCR or other amplification technique prior to analysis. RNA or cDNA may also be used in the same ways. Using amplification, characterization of the species and strain of prokaryote present in an individual, may be made by an analysis of the genotype of the prokaryote gene. Deletions and insertions can be detected by a change in size of the amplified product in comparison to the genotype of a reference sequence. Point mutations can be identified by hybridizing amplified DNA to labeled polynucleotide sequences of the invention. Perfectly matched sequences can be distinguished from mismatched duplexes by RNase digestion or by differences in melting temperatures. DNA sequence differences may also be detected by alterations in the electrophoretic mobility of the DNA fragments in gels, with or without denaturing agents, or by direct DNA sequencing. See, *e.g.*, Myers et al., *Science*, 230: 1242 (1985). Sequence changes at specific locations also may be revealed by nuclease protection assays, such as RNase and S1 protection or a chemical cleavage method. See, *e.g.*, Cotton et al., *Proc. Natl. Acad. Sci., USA*, 85: 4397-4401

(1985).

Cells carrying mutations or polymorphisms in the gene of the invention may also be detected at the DNA level by a variety of techniques, to allow for serotyping, for example. For example, RT-PCR can be used to detect mutations. It is particularly preferred to use RT-PCR in conjunction with automated detection systems, such as, for example, GeneScan. RNA or cDNA may also be used for the same purpose, PCR or RT-PCR. As an example, PCR primers complementary to a nucleic acid encoding a polypeptide of the invention can be used to identify and analyze mutations. These primers may be used for, among other things, amplifying a DNA of the invention isolated from a sample derived from an individual. The primers may be used to amplify the gene isolated from an infected individual such that the gene may then be subject to various techniques for elucidation of the DNA sequence. In this way, mutations in the DNA sequence may be detected and used to diagnose infection and to serotype and/or classify the infectious agent.

The invention further provides a process for diagnosing disease, preferably bacterial infections, more preferably infections by *Streptococcus pneumoniae*, and most preferably disease, comprising determining from a sample derived from an individual a increased level of expression of polynucleotide having the sequence of Table 1. Increased or decreased expression of a polynucleotide of the invention can be measured using any one of the methods well known in the art for the quantitation of polynucleotides, such as, for example, amplification, PCR, RT-PCR, RNase protection, Northern blotting and other hybridization methods.

In addition, a diagnostic assay in accordance with the invention for detecting over-expression of a polypeptide of the invention compared to normal control tissue samples may be used to detect the presence of an infection, for example. Assay techniques that can be used to determine levels of a protein, in a sample derived from a host are well-known to those of skill in the art. Such assay methods include radioimmunoassays, competitive-binding assays, Western Blot analysis and ELISA assays.

Antibodies

The polypeptides of the invention or variants thereof, or cells expressing them can be used as an immunogen to produce antibodies immunospecific for such polypeptides. "Antibodies" as used herein includes monoclonal and polyclonal antibodies, chimeric, single chain, simianized antibodies and humanized antibodies, as well as Fab fragments, including the products of an Fab immunoglobulin expression library.

Antibodies generated against the polypeptides of the invention can be obtained by administering the polypeptides or epitope-bearing fragments, analogues or cells to an animal, preferably a nonhuman, using routine protocols. For preparation of monoclonal antibodies, any technique known in the art that provides antibodies produced by continuous cell line cultures can be used. Examples include various techniques, such as those in Kohler, G. and Milstein, C., *Nature* 256: 495-497 (1975); Kozbor *et al.*, *Immunology Today* 4: 72 (1983); Cole *et al.*, pg. 77-96 in *MONOCLONAL ANTIBODIES AND CANCER THERAPY*, Alan R. Liss, Inc. (1985).

Techniques for the production of single chain antibodies (U.S. Patent No. 4,946,778) can be adapted to produce single chain antibodies to polypeptides of this invention. Also, transgenic mice, or other organisms such as other mammals, may be used to express humanized antibodies.

Alternatively phage display technology may be utilized to select antibody genes with binding activities towards the polypeptide either from repertoires of PCR amplified v-genes of lymphocytes from humans screened for possessing recognition of a polypeptide of the invention or from naive libraries (McCafferty, J. *et al.*, (1990), *Nature* **348**, 552-554; Marks, J. *et al.*, (1992) *Biotechnology* **10**, 779-783). The affinity of these antibodies can also be improved by chain shuffling (Clackson, T. *et al.*, (1991) *Nature* **352**, 624-628).

If two antigen binding domains are present each domain may be directed against a different epitope - termed 'bispecific' antibodies.

The above-described antibodies may be employed to isolate or to identify clones expressing the polypeptides to purify the polypeptides by affinity chromatography.

Thus, among others, antibodies against a polypeptide of the invention may be employed to treat disease.

Polypeptide variants include antigenically, epitopically or immunologically equivalent variants that form a particular aspect of this invention. The term "antigenically equivalent derivative" as used herein encompasses a polypeptide or its equivalent which will be specifically recognized by certain antibodies which, when raised to the protein or polypeptide according to the invention, interfere with the immediate physical interaction between pathogen and mammalian host. The term "immunologically equivalent derivative" as used herein encompasses a peptide or its equivalent which when used in a suitable formulation to raise antibodies in a vertebrate, the antibodies act to interfere with the immediate physical interaction between pathogen and mammalian host.

The polypeptide, such as an antigenically or immunologically equivalent derivative or a fusion protein thereof is used as an antigen to immunize a mouse or other animal such as a rat or chicken. The fusion protein may provide stability to the polypeptide. The antigen may be associated, for example by conjugation, with an immunogenic carrier protein for example bovine serum albumin (BSA) or keyhole limpet haemocyanin (KLH). Alternatively a multiple antigenic peptide comprising multiple copies of the protein or polypeptide, or an antigenically or immunologically equivalent polypeptide thereof may be sufficiently antigenic to improve immunogenicity so as to obviate the use of a carrier.

Preferably, the antibody or variant thereof is modified to make it less immunogenic in the individual. For example, if the individual is human the antibody may most preferably be "humanized"; where the complementarity determining region(s) of the hybridoma-derived antibody has been transplanted into a human monoclonal antibody, for example as described in Jones, P. et al. (1986), *Nature* 321, 522-525 or Tempest et al., (1991) *Biotechnology* 9, 266-273.

The use of a polynucleotide of the invention in genetic immunization will preferably employ a suitable delivery method such as direct injection of plasmid DNA into muscles (Wolff et al., *Hum Mol Genet* 1992, 1:363, Manthorpe et al., *Hum. Gene Ther.* 1993:4, 419), delivery of DNA complexed with specific protein carriers (Wu et al., *J Biol Chem.* 1989: 264,16985), coprecipitation of DNA with calcium phosphate (Benvenisty & Reshef, *PNAS*, 1986:83,9551), encapsulation of DNA in various forms of liposomes (Kaneda et al., *Science* 1989:243,375), particle bombardment (Tang et al., *Nature* 1992, 356:152, Eisenbraun et al., *DNA Cell Biol* 1993, 12:791) and *in vivo* infection using cloned retroviral vectors (Seeger et al., *PNAS* 1984:81,5849).

Antagonists and agonists - assays and molecules

Polypeptides of the invention may also be used to assess the binding of small molecule substrates and ligands in, for example, cells, cell-free preparations, chemical libraries, and natural product mixtures. These substrates and ligands may be natural substrates and ligands or may be structural or functional mimetics. See, *e.g.*, Coligan *et al.*, *Current Protocols in Immunology* 1(2): Chapter 5 (1991).

The invention also provides a method of screening compounds to identify those which enhance (agonist) or block (antagonist) the action of a polypeptides or polynucleotides of the invention, particularly those compounds that are bacteriostatic and/or bacteriocidal. The method of screening may involve high-throughput techniques. For example, to screen for

agonists or antagonists, a synthetic reaction mix, a cellular compartment, such as a membrane, cell envelope or cell wall, or a preparation of any thereof, comprising a polypeptide of the invention and a labeled substrate or ligand of such polypeptide is incubated in the absence or the presence of a candidate molecule that may be an agonist or antagonist of a polypeptide of the invention. The ability of the candidate molecule to agonize or antagonize a polypeptide of the invention is reflected in decreased binding of the labeled ligand or decreased production of product from such substrate. Molecules that bind gratuitously, *i.e.*, without inducing the effects of a polypeptide of the invention are most likely to be good antagonists. Molecules that bind well and increase the rate of product production from substrate are agonists. Detection of the rate or level of production of product from substrate may be enhanced by using a reporter system. Reporter systems that may be useful in this regard include but are not limited to colorimetric labeled substrate converted into product, a reporter gene that is responsive to changes in polynucleotide or polypeptide activity, and binding assays known in the art.

Another example of an assay for antagonists of polypeptides of the invention is a competitive assay that combines any such polypeptide and a potential antagonist with a compound which binds such polypeptide, natural substrates or ligands, or substrate or ligand mimetics, under appropriate conditions for a competitive inhibition assay. A polypeptide of the invention can be labeled, such as by radioactivity or a colorimetric compound, such that the number of such polypeptide molecules bound to a binding molecule or converted to product can be determined accurately to assess the effectiveness of the potential antagonist.

Potential antagonists include small organic molecules, peptides, polypeptides and antibodies that bind to a polynucleotide or polypeptide of the invention and thereby inhibit or extinguish its activity. Potential antagonists also may be small organic molecules, a peptide, a polypeptide such as a closely related protein or antibody that binds the same sites on a binding molecule, such as a binding molecule, without inducing activities induced by a polypeptide of the invention, thereby preventing the action of such polypeptide by excluding it from binding.

Potential antagonists include a small molecule that binds to and occupies the binding site of the polypeptide thereby preventing binding to cellular binding molecules, such that normal biological activity is prevented. Examples of small molecules include but are not limited to small organic molecules, peptides or peptide-like molecules. Other potential antagonists include antisense molecules (see Okano, *J. Neurochem.* 56: 560 (1991); *OLIGODEOXYNUCLEOTIDES AS ANTISENSE INHIBITORS OF GENE EXPRESSION*,

CRC Press, Boca Raton, FL (1988), for a description of these molecules). Preferred potential antagonists include compounds related to and variants of a polypeptide of the invention.

Each of the DNA sequences provided herein may be used in the discovery and development of antibacterial compounds. The encoded protein, upon expression, can be used as a target for the screening of antibacterial drugs. Additionally, the DNA sequences encoding the amino terminal regions of the encoded protein or Shine-Delgarno or other translation facilitating sequences of the respective mRNA can be used to construct antisense sequences to control the expression of the coding sequence of interest.

The invention also provides the use of the polypeptide, polynucleotide or inhibitor of the invention to interfere with the initial physical interaction between a pathogen and mammalian host responsible for sequelae of infection. In particular the molecules of the invention may be used: in the prevention of adhesion of bacteria, in particular gram positive bacteria, to mammalian extracellular matrix proteins on in-dwelling devices or to extracellular matrix proteins in wounds; to block protein-mediated mammalian cell invasion by, for example, initiating phosphorylation of mammalian tyrosine kinases (Rosenshine *et al.*, *Infect. Immun.* 60:2211 (1992); to block bacterial adhesion between mammalian extracellular matrix proteins and bacterial proteins that mediate tissue damage and; to block the normal progression of pathogenesis in infections initiated other than by the implantation of in-dwelling devices or by other surgical techniques.

The antagonists and agonists of the invention may be employed, for instance, to inhibit and treat disease.

Helicobacter pylori (herein *H. pylori*) bacteria infect the stomachs of over one-third of the world's population causing stomach cancer, ulcers, and gastritis (International Agency for Research on Cancer (1994) Schistosomes, Liver Flukes and *Helicobacter Pylori* (International Agency for Research on Cancer, Lyon, France; <http://www.uicc.ch/ecp/ecp2904.htm>). Moreover, the international Agency for Research on Cancer recently recognized a cause-and-effect relationship between *H. pylori* and gastric adenocarcinoma, classifying the bacterium as a Group I (definite) carcinogen. Preferred antimicrobial compounds of the invention found using screens provided by the invention, particularly broad-spectrum antibiotics, should be useful in the treatment of *H. pylori* infection. Such treatment should decrease the advent of *H. pylori*-induced cancers, such as gastrointestinal carcinoma. Such treatment should also cure gastric ulcers and gastritis.

Vaccines

Another aspect of the invention relates to a method for inducing an immunological response in an individual, particularly a mammal which comprises inoculating the individual with a polypeptide of the invention, or a fragment or variant thereof, adequate to produce antibody and/ or T cell immune response to protect said individual from infection, particularly bacterial infection and most particularly *Streptococcus pneumoniae* infection. Also provided are methods whereby such immunological response slows bacterial replication. Yet another aspect of the invention relates to a method of inducing immunological response in an individual which comprises delivering to such individual a nucleic acid vector to direct expression of a polynucleotide or polypeptide of the invention, or a fragment or a variant thereof, for expressing such polynucleotide or polypeptide, or a fragment or a variant thereof *in vivo* in order to induce an immunological response, such as, to produce antibody and/ or T cell immune response, including, for example, cytokine-producing T cells or cytotoxic T cells, to protect said individual from disease, whether that disease is already established within the individual or not. One way of administering the gene is by accelerating it into the desired cells as a coating on particles or otherwise. Such nucleic acid vector may comprise DNA, RNA, a modified nucleic acid, or a DNA/RNA hybrid.

A further aspect of the invention relates to an immunological composition which, when introduced into an individual capable or having induced within it an immunological response, induces an immunological response in such individual to a polynucleotide of the invention or protein coded therefrom, wherein the composition comprises a recombinant polynucleotide or protein coded therefrom comprising DNA which codes for and expresses an antigen of said polynucleotide or protein coded therefrom. The immunological response may be used therapeutically or prophylactically and may take the form of antibody immunity or cellular immunity such as that arising from CTL or CD4+ T cells.

A polypeptide of the invention or a fragment thereof may be fused with co-protein which may not by itself produce antibodies, but is capable of stabilizing the first protein and producing a fused protein which will have immunogenic and protective properties. Thus fused recombinant protein, preferably further comprises an antigenic co-protein, such as lipoprotein D from *Hemophilus influenzae*, Glutathione-S-transferase (GST) or beta-galactosidase, relatively large co-proteins which solubilize the protein and facilitate production and purification thereof. Moreover, the co-protein may act as an adjuvant in the

sense of providing a generalized stimulation of the immune system. The co-protein may be attached to either the amino or carboxy terminus of the first protein.

Provided by this invention are compositions, particularly vaccine compositions, and methods comprising the polypeptides or polynucleotides of the invention and immunostimulatory DNA sequences, such as those described in Sato, Y. *et al.* Science 273: 352 (1996).

Also, provided by this invention are methods using the described polynucleotide or particular fragments thereof which have been shown to encode non-variable regions of bacterial cell surface proteins in DNA constructs used in such genetic immunization experiments in animal models of infection with *Streptococcus pneumoniae* will be particularly useful for identifying protein epitopes able to provoke a prophylactic or therapeutic immune response. It is believed that this approach will allow for the subsequent preparation of monoclonal antibodies of particular value from the requisite organ of the animal successfully resisting or clearing infection for the development of prophylactic agents or therapeutic treatments of bacterial infection, particularly *Streptococcus pneumoniae* infection, in mammals, particularly humans.

The polypeptide may be used as an antigen for vaccination of a host to produce specific antibodies which protect against invasion of bacteria, for example by blocking adherence of bacteria to damaged tissue. Examples of tissue damage include wounds in skin or connective tissue caused, e.g., by mechanical, chemical or thermal damage or by implantation of indwelling devices, or wounds in the mucous membranes, such as the mouth, mammary glands, urethra or vagina.

The invention also includes a vaccine formulation which comprises an immunogenic recombinant protein of the invention together with a suitable carrier. Since the protein may be broken down in the stomach, it is preferably administered parenterally, including, for example, administration that is subcutaneous, intramuscular, intravenous, or intradermal. Formulations suitable for parenteral administration include aqueous and non-aqueous sterile injection solutions which may contain anti-oxidants, buffers, bacteriostats and solutes which render the formulation isotonic with the bodily fluid, preferably the blood, of the individual; and aqueous and non-aqueous sterile suspensions which may include suspending agents or thickening agents. The formulations may be presented in unit-dose or multi-dose containers, for example, sealed ampules and vials and may be stored in a freeze-dried condition requiring only the addition of the sterile liquid carrier

immediately prior to use. The vaccine formulation may also include adjuvant systems for enhancing the immunogenicity of the formulation, such as oil-in water systems and other systems known in the art. The dosage will depend on the specific activity of the vaccine and can be readily determined by routine experimentation.

While the invention has been described with reference to certain protein, such as, for example, those set forth in Table 1, it is to be understood that this covers fragments of the naturally occurring protein and similar proteins with additions, deletions or substitutions which do not substantially affect the immunogenic properties of the recombinant protein.

Compositions, kits and administration

The invention also relates to compositions comprising the polynucleotide or the polypeptides discussed above or their agonists or antagonists. The polypeptides of the invention may be employed in combination with a non-sterile or sterile carrier or carriers for use with cells, tissues or organisms, such as a pharmaceutical carrier suitable for administration to a subject. Such compositions comprise, for instance, a media additive or a therapeutically effective amount of a polypeptide of the invention and a pharmaceutically acceptable carrier or excipient. Such carriers may include, but are not limited to, saline, buffered saline, dextrose, water, glycerol, ethanol and combinations thereof. The formulation should suit the mode of administration. The invention further relates to diagnostic and pharmaceutical packs and kits comprising one or more containers filled with one or more of the ingredients of the aforementioned compositions of the invention.

Polypeptides and other compounds of the invention may be employed alone or in conjunction with other compounds, such as therapeutic compounds.

The pharmaceutical compositions may be administered in any effective, convenient manner including, for instance, administration by topical, oral, anal, vaginal, intravenous, intraperitoneal, intramuscular, subcutaneous, intranasal or intradermal routes among others.

In therapy or as a prophylactic, the active agent may be administered to an individual as an injectable composition, for example as a sterile aqueous dispersion, preferably isotonic.

Alternatively the composition may be formulated for topical application for example in the form of ointments, creams, lotions, eye ointments, eye drops, ear drops, mouthwash, impregnated dressings and sutures and aerosols, and may contain appropriate conventional additives, including, for example, preservatives, solvents to assist drug

penetration, and emollients in ointments and creams. Such topical formulations may also contain compatible conventional carriers, for example cream or ointment bases, and ethanol or oleyl alcohol for lotions. Such carriers may constitute from about 1% to about 98% by weight of the formulation; more usually they will constitute up to about 80% by weight of the formulation.

For administration to mammals, and particularly humans, it is expected that the daily dosage level of the active agent will be from 0.01 mg/kg to 10 mg/kg, typically around 1 mg/kg. The physician in any event will determine the actual dosage which will be most suitable for an individual and will vary with the age, weight and response of the particular individual. The above dosages are exemplary of the average case. There can, of course, be individual instances where higher or lower dosage ranges are merited, and such are within the scope of this invention.

In-dwelling devices include surgical implants, prosthetic devices and catheters, i.e., devices that are introduced to the body of an individual and remain in position for an extended time. Such devices include, for example, artificial joints, heart valves, pacemakers, vascular grafts, vascular catheters, cerebrospinal fluid shunts, urinary catheters, continuous ambulatory peritoneal dialysis (CAPD) catheters.

The composition of the invention may be administered by injection to achieve a systemic effect against relevant bacteria shortly before insertion of an in-dwelling device. Treatment may be continued after surgery during the in-body time of the device. In addition, the composition could also be used to broaden perioperative cover for any surgical technique to prevent bacterial wound infections, especially *Streptococcus pneumoniae* wound infections.

Many orthopedic surgeons consider that humans with prosthetic joints should be considered for antibiotic prophylaxis before dental treatment that could produce a bacteremia. Late deep infection is a serious complication sometimes leading to loss of the prosthetic joint and is accompanied by significant morbidity and mortality. It may therefore be possible to extend the use of the active agent as a replacement for prophylactic antibiotics in this situation.

In addition to the therapy described above, the compositions of this invention may be used generally as a wound treatment agent to prevent adhesion of bacteria to matrix proteins exposed in wound tissue and for prophylactic use in dental treatment as an alternative to, or in conjunction with, antibiotic prophylaxis.

Alternatively, the composition of the invention may be used to bathe an indwelling device immediately before insertion. The active agent will preferably be present at a concentration of 1µg/ml to 10mg/ml for bathing of wounds or indwelling devices.

A vaccine composition is conveniently in injectable form. Conventional adjuvants may be employed to enhance the immune response. A suitable unit dose for vaccination is 0.5-5 microgram/kg of antigen, and such dose is preferably administered 1-3 times and with an interval of 1-3 weeks. With the indicated dose range, no adverse toxicological effects will be observed with the compounds of the invention which would preclude their administration to suitable individuals.

Each reference disclosed herein is incorporated by reference herein in its entirety. Any patent application to which this application claims priority is also incorporated by reference herein in its entirety.

TABLES

Certain pertinent data for preferred polypeptide and polynucleotide embodiments of the invention are summarized in Tables 1 and 2.

Provided in Table 1 are sequence search results providing characterization information regarding certain preferred polynucleotides (denoted as "Assembly") and polypeptides of the invention encoded thereby. For each polynucleotide in Table 1, there is listed the closest homologue of each polypeptide encoded by each ORF in such polynucleotide. This determination of homology is based on a comparison of the sequences of in Table 1 with sequences available in the public domain (see heading entitled "Description" for the homologue name). Where no significant homologue was detected the term "unknown" appears after the heading "Description". Preferred polypeptides encoded by the ORFs of the invention, particularly full length proteins either obtained using such ORFs or encoded entirely by such ORFs, are ones that have a biological function of the homologue listed, among other functions. The analysis used to determine each homologue listed in Table 1 was either BlastP and/or BlastX and/or MPSearch, each of which is well known. Also provided in Table 1 is the amino acid sequence encoded by each ORF. An "Assembly ID" number provides a convenient way to correlate the polynucleotide sequence with the ORF or ORFs it comprises and the polypeptides encoded by these ORFs, as well as to correlate such sequences with other pertinent information provided in Tables 1 and 2. Following the heading "ORF Predictions" the nucleotides at the beginning and end of the ORF sequence are set forth ("Start" and "End" respectively). The direction of translation

on the polynucleotide depicted is denoted by an "F" for forward or an "R" for reverse (reverse being translated on the opposite strand from the one depicted). The length of each amino acid sequence is also indicated in a column entitled "Length." Below these data is shown the amino acid sequence encoded by the ORF. If a given polynucleotide comprises one ORF, then in the column entitled "ORF #" there is the numeral one. If it encodes two, there are the numerals one and two in the column, and so on.

TABLE 1

Assembly ID: 3049156

Assembly Length: 495bp

```
>[SEQ ID NO:1] 3049156 Strep Assembly -- Assembly id#3049156
CTCGGTGATAGAAATAGTGTAATCATGCTTTTCTCTTCTTATCTATACTTTGCTACTTCT
ATTATACAAAAAATAAAGCGCTTGACTAGGGATTTTGTAGAAAAAAGCCTATTTTTTCA
AGAAAAATAGGCTTTTTGCGAACGATTGACACAATTGGATTTGGTTAATTCACCTTAAC
GATGGTTTTAAACGATATATATTTTTATATATGTAAATTAAAACTTCTTTCCTTTCAC
TCCTACGACTTTTCAGATACAGATAGCCAAAGAAGTTTTCATAGAGGGCAAAAAGAGGA
GGAAGGCATGAAGAAAGAAGGTCTCTGGCAAAATCATAATAACAGGATCCTTGGCTGGAT
CAAAAAGCCAGGTATCATCTCCACAAAGAGAATTTGATGGAAAAGAGTAAAGAATTGGT
CAAAACCAATCAAACTCCCCCAAGTCCATCATCACAGGTAAGACTACTAGAGCCAGGAG
ACTTTTTTCGATAAAG
```

ORF Predictions:

ORF #	Start	End	Direction	Length
-----	-----	-----	-----	-----
1	236	385	R	50 aa

```
>[SEQ ID NO:88] 3049156-1 ORF translation from 236-385, direction R
VGDDTWLFDPAKDPVIMILPETFFLHAFLLFFALYENFFGYLYLKSRRK*
```

Description:

unknown

Assembly ID: 3049862

Assembly Length: 529bp

```
>[SEQ ID NO:2] 3049862 Strep Assembly -- Assembly id#3049862
CTAGAGCAAGTATTTTTCAAACCTTTTCCGAATAAATAGATAGAGCCAGAGAATTTAGTA
AACCTAGATTTAAAAATGTGCTATAACATAATATATTGAATCTATAATAGTACACCTTGA
CTGCTAAATATTTCTATAAATTAATTTGACTTTCCTGATAGAGTTATTCACATCTTATT
TCAACTCACTATAGAAGGAGGAATAGGAGGATTCTCAGACATCCGGGCATCAGCCCAACT
```

AATGATTTGATTGCTAAGAAAATATTCAGCAATCCAGAAATCACTTGTCAATTTATTCGC
 GATATGCTGGACTTGCCAGCAAAAAATGTTGACCATTTTGGAGGGAAGCGATATTCACGT
 ATTACTCTCCATGCCTTACTCAGTGCAGGATTTTATACCAGTATAGACGTCTTGCGGA
 GTTGATAACGGTACTCAAGTAATTATGAGATTCAAGTCCATCATCAGAATTTTTCATC
 AATCACTTGTGGACTTACCTGTGCAGTCAGGTTAATCAAATCTTGAAAA

ORF Predictions:

ORF #	Start	End	Direction	Length
1	383	526	F	48 aa

>[SEQ ID NO:89] 3049862-1 ORF translation from 383-526, direction F
 VQDFYTSIDVLAELDNGTQVIIEIQVHHQNFSSITCGLTCAVRLIKS*

Description:

unknown

Assembly ID: 3112810

Assembly Length: 885bp

>[SEQ ID NO:3] 3112810 Strep Assembly -- Assembly id#3112810
 CTCATCATCTGTCAAAAAGCGTTTCTTAGCAGTCGTGATATCCATAAAATAATCTAATAT
 CACGATTTCTCATCCGCAAAGAAAGGAAGGCTGACCAACTCCAGTGCCACATCCTTGTA
 AACTACTTCTTGCAATATCAAAGTAGGCAAAGTTGAGGTCAGCAGAATCATACCCAATCTG
 TTTCAACACTTGACTCTTCATCACTTCAAAGTACCTGATCTGTCCCTGTAAATAGGCG
 CAGGCTCGGTAAATTCGATAAAGTCAACTTCTGACTTTCTTCAATGGCTAGCATCGTCTC
 TCCTTTCTTCAGATTTTTCGATTTAATTTAGTCAATATAGCGCAATTTCCCACGGAAATC
 TTCTAAGCTCTCGTAGCCTTTTCCACCATGATTGCTTTTCAGTTCATTGGTAAAGCGGTC
 AAAAGCACTGACGCCTTCTTTGTGAAGGGTCGTTCCCACCTGCACCATACTTGCTCCACA
 GAGGATGTGTTCAAAGGCATCTCGACCAAGTCAGAACGCCACCTGTTCCGATAATTTGGAT
 TTGAGGATTTAAACGTTGATAAAAGGCGTGAACATTGGCTAGAGCAGTCGGTTTGATGTA
 TTATCCACCAATTCCACCAAACCATTCTTAGGCCGAATAACGACAGATTTCGTCTTCTAT
 ATAGAGGCCGTTTCCGATAGAGTTAACGCAGTTGACAACTTGAGCGGATATTTGTTGAA
 AATAGCTGCCGCTTGATCAAAGTGAACAATATCAAATAAGGTGGCAATTTAATTCCAAG
 AGGTTTGGTGAAGTAAGCAAACACTTCTGCCAAATCCGGTCTGTTGTCTCAAATCATA
 GGCAATCTGAGGTTTACCTGGAACATTTGGACAGGAAAGATTTAG

ORF Predictions:

ORF #	Start	End	Direction	Length
1	601	804	R	68 aa

>[SEQ ID NO:90] 3112810-2 ORF translation from 601-804, direction R
 35

VFAYFTKPLGIKLPPYFDIVHFDQAAAI FNKYPLKFVNCVNSIGNGLYIEDES SVVIRPKN
GFGGIGG*

Description:

LLCPYRDA NCBI gi: 511014 - *Lactococcus lactis*. DIHYDROOROTATE
DEHYDROGENASE (EC 1.3.3.1) (DIHYDROOROTATE OXIDASE)

Assembly ID: 3112866

Assembly Length: 925bp

>[SEQ ID NO:4] 3112866 Strep Assembly -- Assembly id#3112866
TCTTGGCCAACTGCATGGAGTTCAGCGGTCAATTTCAACGCACCTGAGAAACAGACCCCT
GCACCCCTGAAATCTCAGGAGACATGATGGTCTGGATGGAATCAATAATGAGAAAGTCTG
GCTGGATACGCTACCACTTCTGCACGAACACTCTGCATATTGGTCTCTGCATAGAGATAA
AACTCACTATCAAAATCACCTAAGCGCTCTGCACGTAGTTTAATCTGCTGGGCAGACTCC
TCCCCACTGACATAGAGAACTGTCCCCACTTGGGACAACCTGGGTGAGACTTGTAGGAGA
AGAGTTGATTTCCCAATCCCAGGATCCCCACCGATGAGGACGAGACTTTCCTGGTACAAC
TCCGCCTCCAAGCACACGGTTGAATTCCTCCATCTCCGTCTTGGTTCGATTGACATTGAT
GGAAGTCACCTCAGCTAGTTTCATGGGCTTGGTTTTCTCACCTGTCAAGGACACACGCGC
ATTCTTGACCTCGGCAACCTCAACCTCTTCCACAAAAGAAGACCAAGACCCACAGTTGGG
GCAACGTCCCAGATATTTAGGGGAATTATACCCACAATTTTGACATACAAATGTCGCTTT
TTTCTTTGCGATGACAAACCTCTTTCTATATCTCTAACTCACACTCAATCACTTGGCAAA
AATCAATCTTCTCATTTGGCACAAACTGGCGCATGAGCATTTCGATGAGCAACAACTACCA
CAGTCTGATGTTCTCGATACTTAGACATACATTCTAGAAACCGAGACTTCATTTCCGTAG
CTGTCTCATATTGAATAGGACTATTAGGAAGCAACTCCCCCTTGTTTTCTAAAAACAGTC
TTCTAGCTGTTTTCAAAGTTTTCTATTCTGTTTTATAGACCTGCCATTCATGTAATAAAG
GCTCTACTCTTAAAGGAAGACCCGT

ORF Predictions:

ORF #	Start	End	Direction	Length
1	220	513	R	98 aa

>[SEQ ID NO:91] 3112866-2 ORF translation from 220-513, direction R
VEEVEVAEVKNARVSLTGEKTKPMKLAEVT SINVNRTKTEMEEFNRLGGGVVPGKSRPH
RWGSWDWEINSSPTSLNPVPSGDSSLCQWGGVCPAD*

Description:

SMS PROTEIN. - *ESCHERICHIA COLI*.

Assembly ID: 3113664

Assembly Length: 602bp

>[SEQ ID NO:5] 3113664 Strep Assembly -- Assembly id#3113664
 TTATGTCAGTGGGATTACGCCTAATCTCCCAGAAGCAGAATTATTATCCGGTCAGGAAAT
 TAAACCTTGGNAGACATGAAAACCTGCAGCGCAGAAATTGCATGATTTAGGAGCGCCAGC
 AGTCATTATCAAAGGGAGGCAATCGTCTTAGTCAGGACAAGGCTGTGGATGTCTTTTATG
 ATGGACAGACCTTTACTATCCTAGAAAATCCAGTTATCCAAGGCCAAAATGCTGGTGCAG
 GTTGACCTTTGCCTCTAGCATTGCCAGTCACTTGGTTAAAGGTGATAAACTTTTGCCAG
 CAGTAGAAAGCTCTAAGGCTTTCGTTTATCGTGCTATTGCACAAGCAGATCAGTATGGAG
 TAAGACAATATGAAGCAAACAAAACAATAAAATCGCCCTTGTATCCCTATTAACCGCC
 CTTTCTGTGGTTCTAGGTTATTTCTTAAAAATCCCAACACCTACAGGNATTCTAACTCTT
 TTAGATGCTGGTGTCTTCTTTGCGGCCTTTTACTTTGGTAGTCGTGAAGGAGCGGTAGTC
 GGAGGACTAGCAAGTTTCTTGCTTGACCTCTTATCAGGCTACCCTCAGTGGATGTTTTTT
 AG

ORF Predictions:

ORF #	Start	End	Direction	Length
1	165	392	F	76 aa

>[SEQ ID NO:92] 3113664-1 ORF translation from 165-392, direction F
 VDVFYDQGTFITILENPVIQGNAGAGCTFASSIASHLVKGDKLLPAVESSKAFVYRAIAQ
 ADQYGVVRQYEANKNN*

Description:

Thi protein - Rhizobium meliloti

Assembly ID: 3113716

Assembly Length: 456bp

>[SEQ ID NO:6] 3113716 Strep Assembly -- Assembly id#3113716
 CTGGATACTAAGAGAAATCAAAAAAGCACTCTAGGATAGAGGCCTAAAGTGCTTAGTTTC
 AAGGCTTTACAGCCTATCATATTTAATAAAATATTACAACATCTTGTTGTAGAATTCAAC
 GACAAGTGCTTCGTTGATTTCTGGGTTGATTTTCGTCGCGTTCTGGCAAGCGAGTCAATGA
 ACCTTCCAATTTTTCAGCGTCGAATGATACGAATGCTGGACGTCCAAGAGTAGCTTCTAC
 TGCTTCAAGGATTGCTGGAACCTTCAATGATTTTTCACGAACTGAGATCACTTGACCTGC
 AGTTACGCGGTATGATGGGATATCAACGCGTTTCCCGTCAACAAGGATGTGACCGCTGGT
 TTACAAATTGGACCAAACCTTGACGACCAGTAGTCGCGAGACCAAGACGGTAAACAACGTT
 ATCCAAACGACGTTCCAAAAGAAGCATAAAGTTGAA

ORF Predictions:

ORF #	Start	End	Direction	Length
1	94	291	R	66 aa

>[SEQ ID NO:93] 3113716-1 ORF translation from 94-291, direction R
VISVREKSLKVPAILLEAVEATLGRPAFVSFDAEKLEGLSLTRLPERDEINPEINEALVVEF
YNKML*

Description:

30S RIBOSOMAL PROTEIN S4 (BS4). - BACILLUS SUBTILIS.

Assembly ID: 3174176

Assembly Length: 1961bp

>[SEQ ID NO:7] 3174176 Strep Assembly -- Assembly id#3174176
CTAATATAGAATAATCACCGCCGTTGTGAAAGAACGATTGGATGATAATCCAATCGTTCA
GGGAAATTGGAAGACCTTGGGTTTCCAATTTAGGCATGAGACACCTTTGGTGGCTGCTGC
CGTCCCTCACAAAGCTAAGGTGATTGTTGAAAAAGAGGAAAAAGGAGAAGAAATGAAACCA
GTAATTTCCATCATCATGGGCTCAAAATCCGACTGGGCAACCATGCAAAAAACAGCAGAA
GTCCTAGACCGCTTCGGTGTAGCCTACGAAAAGAAAGTTGTTTCCGCACACCGTACACCA
GACCTCATGTTCAAACATGCAGAAGAAGCCCGTAGTCGTGGCATCAAGATCATCATCGCA
GGTGTGGTGGCGCAGCGCATTTGCCAGGCATGGTAGCTGCCAAAACAACCTTCCAGTC
ATTGGTGTGCCAGTCAAGTCTCGTGCTCTTAGTGGAGTGGATTCACTCTATTCTATCGTT
CAGATGCCGGGTGGGGTGCCTGTTGCGACCATGGCTATCGGTGAACTCTTTTTTTAGGATA
TAAACAGGGTTCGGATAAGTTTTTTTTGCAAGGTGGATGATGGCTACATTGTAATGTTTT
CCTTGTTCTAACTTAGTCTTAAAGCAGGTGAAAAGTGAGGGCATGCTTTGGCAGCTTGT
ATGAGTACCTACCGCAGATAAGGGGAACCCCGTTTGACCATCCTCCCAGCTAAATCAATC
TGACCTGACTGATAAATAGAAGAATCCAGTCCAGCGAAAGCTTGTAATTGAGCAGGATTA
TCAAAGGCATGAATATTTTGAATCTCGGCTAAAATGACCGCCCCCTAAACGATTCTCAATC
CCAGTAACCGTCGTGATGACCGAGTTTAACTCAGCCATCAAGTCATTGACACATTTTTCC
GCCTTGTCATGAGCCTCTTGTAATGTTTGATGTTTTTATTACACGAGATAAAACGTCTA
TGCGTTATCAAACCTATTACCAATTAAAACAAATGTGGTTAGATCCTTTCGGAAATTGTC
AAGCGATTGGAGGAAATGAACTAATCCACAGCGGCTTATTCCAAGTATACCACTTGGGCT
TTGGCAGTAGCTAACTGCGCTAAATATAATATAAGGAGGAGTAAAATGAAGACAGTTCAA
TTTTTTTGGCATTATTTTAAGGTCTACAAGTTCTCATTTGTAGTTGTCATCCTGATGATT
GTTCTGGCGACTTTTGCCCCAAGCCCTCTTTCCAGTCTTTTCTGGACAAGCGGTGACGCAG
CTAGCCAATTTAGTTCAAGCTTATCAAAATGGGCAATCCAGAACTTGATGGCAAAGCCT
ATCAGGAATTCATGGTCAATCTTGGCCTGCTGGTTTTGGGTTCTATTTATCTCTAGGTGT
AATATAAACATGTGTCTCATGACGCGCGTGATTGCAGAATCGACCAACGAGATGCGCAAA
GGTCTCTTTGGTAAGCTTGCTCAGTTGACGGTTTCTTTCTTTGACCGTCGACAAGATGGC
GATATCCTGTCTCATTTTACCAGTGATTTGGATAATATCCTCCAAGCCTTTAACGAAAGC
TTGATTCAAGTTCATGAGCAATATTGTTTTATACATTGGTCTGATTCTTGTCATGTTTTCG
AGAAATGTGACGCTGGCTCTCATCACCATTGCCAGCACCCCATTTGGCTTTTCTTATGCTG
ATTTTTCATCGTGAAAATGGCACGTAAATACACCAACCTCCAGCAGAAAGAGGTAGGGAAG
CTCAACGCCTATATGGATGAGAGCATCTCAGGCCAAAAAGCCGTGATTGTGCTAGGAATT
CAAGAGGATATGATGGCAGGATTTCTTGAACAAAATGAGCGCGTGCGCAAGGCAACCTTT
AAAGGAAGAATGTTCTCAGGAATTTCTTTCCCTGTCATGAATGGGATGAGCCTGATTAAT

ACAGCCATCGTCATCTTTGCTGGTTCGGCTGTACTTTTGAA

ORF Predictions:

ORF #	Start	End	Direction	Length
1	139	543	F	135 aa

>[SEQ ID NO:94] 3174176-1 ORF translation from 139-543, direction F
 VIVEKEEEKGEEMKPVISIIMGSKSDWATMQKTAEVLDLRFVAYEKKVVSARHTPDLMFKH
 AEEARSRGIKIIIIAGAGGAHLPGMVAAKTTLPVIGVPVKSRALSGVDSLVSIVQMPGGV
 PVATMAIGELFFRI*

Description:

PHOSPHORIBOSYLAMINOIMIDAZOLE CARBOXYLASE CATALYTIC SUBUNIT (EC
 4.1.1.21) (AIR C ARBOXYLASE) (AIRC). - BACILLUS SUBTILIS.

Assembly ID: 3174186

Assembly Length: 375bp

>[SEQ ID NO:8] 3174186 Strep Assembly -- Assembly id#3174186
 CTATCTCCAAGTNCGNNTTGGAAATNCCTCCGCNANCCACAACATCCAAGCACTTTNCAA
 CGTGNCTGGTCCGGTCCCTCCAGTGCGTCTNACNGCACCTTCAACCTGCNCATGGGTAGG
 TCACATGGCTTCGGGTCTACGTCATGATACTAAGGCGCCCTATTCAGACTCGGNTNCCCT
 AGGGCTCCGTCTCTTCAACTTAACCACGCAACAGAACGTNACCCGCCGGTTCATTCTACA
 AAAGGCAGNCTCTCACCCATTAACGGGCTCGAAGTTGTTGTAGGCACACNGCTTCAGGTN
 CTATTTTACCCCCCTCCCGGGGAGCANCTCAACTGACCCNCACGGCACCGGTGNANNA
 CGGTCACTTAGGGAG

ORF Predictions:

ORF #	Start	End	Direction	Length
1	83	283	F	67 aa

>[SEQ ID NO:95] 3174186-1 ORF translation from 83-283, direction F
 VRXXAPSTCXWVGHMASGLRHDTKAPYSDSXXLGLRLFNLTQQNXTRRFILQKAXSHPL
 TGSNLL*

Description:

unknown

Assembly ID: 3174374

Assembly Length: 665bp

>[SEQ ID NO:9] 3174374 Strep Assembly -- Assembly id#3174374
 GGGGGGGGTNNNTTCTGGGGCCGGGTGNNTCCTNGAAAAAATGCTGGACTTAACGGTTAA
 ATCATTTGAATTGGCCTGTGGATTTTAGCTAGCAATCCAGAGCGAGTTTCTCCAAGACA
 GACCTCTATGAAAAGATCTGGAAAGAACTACGTGGATGACACCAATACCTTGAATGTG
 CATATCCATGCTCTTCGACAGGAGCTGGCAAATATAGTAGTGACCAAACGCCCACTATT
 AAGACAGTTTGGGGGTGGGATATAAGATAGAGAAACCGAGAGGACAAACATGAAACTAA
 AAAGTTATATTTTGGTTGGATATATTATTTCAACCCTCTTAACCATTTTGGTTGTTTTT
 GGGCTGTTCAAAAAATGCTGATTGCGAAAGGCGAGATTTACTTTTTGCTTGGGATGACCA
 TCGTTGCCAGCCTTGTCGGTGCTGGGATTAGTCTCTTTCTCCTATTGCCAGTCTTTACGT
 CGTTGGGCAACTCAAGGAGCATGCCAAGCGGGTAGCGCCAAGGATTTCCCTCCAATTT
 GGANGTTCAAGGTCCCTGTTAAATTTCCCCCATTTAGGGGCAACCTTTTAATGAAANTTT
 CCNTNATTTGCCGGGTANCTTTGAATCCCTNGGAAAAAACCAACNAAAAAAGGGCTTA
 NNCCC

ORF Predictions:

ORF #	Start	End	Direction	Length
1	154	294	F	47 aa

>[SEQ ID NO:96] 3174374-1 ORF translation from 154-294, direction F
 VDDTNTLNVHIHALRQELAKYSSDQTPTIKTWVWGLGYKIEKPRGQT*

Description:

REGULATORY PROTEIN VANR. - ENTEROCOCCUS FAECIUM (STREPTOCOCCUS FAECIUM).

Assembly ID: 3174972

Assembly Length: 989bp

>[SEQ ID NO:10] 3174972 Strep Assembly -- Assembly id#3174972
 CTACGATATCTTTGGTCTTTTGTAAGATATGAGGTCCACCCTTATGCGCCTCAGTTGGCA
 TTTTCATGCGATTCAAGAAGTTGCCCCCTCTTGATCAACCAAACCATACTTGATGTTGGTTC
 CACCGATATCAATTGCAACGTAATATGTCATAAATACCTCCTTTTAGATTAGAGGAAGCG
 CTCCTTGGTTTCACGAATCAAGGCAGCAGCCGCTTCTACAACCTGGACGATCTTCTTCAGT
 CACTGGTGTCAATGGTGAACGAACAGATCCAATATTCAAGCCTTCATTGATTTTCAAGAC
 TTCTTTGATGACACCGTACATATTTCCATGAGCAGAAGTGAGTTTACCAATGATTGCGTT
 GATAGCATACTGCAATTCACGCGCTGTTTCTAGGTCCCTTATCCGCAATCAACTGATTGAG
 TTTCAAGAAGAGTTCTGGCATAGCACCATAAGTACCACCGATAACCAGCCCTAGCCCCCAT
 GAGGCGTCCCTCCTAGGAACTGCTCATCAGGACCATTAAAGACGATATGGTCTTCTCCACC
 AAGGCTGACAAAGGTTTGGATATCTTGAACCTGGCATAGAAGAGTTCTTCACACCGATAAC
 ACGAGGATTTTTCAACATTTCTGTGTAAAGGCTTGGAGTCAAAGCAACCCCTGCCAATTG
 AGGAATGTTGTAAATCACGTAGTCTGTGTTGGAGCTGCAGAACTGATATCGTTCCAGTA
 TTTGGCAACTGAGTTATTCTGGCAAGCGGAAATAAATTGGTGGAATCCGTTGCAATAGCA

TCTACTCCCAAGCTTTCAGCATGGCGAGCAAGTTCCATACTATCTTTAGTATTATTGCAA
GCAACATGGGCAATAATGGTCAATTTACCTTTGGCTACCGCCATGACTTCTTCCAAAATC
AACTTGCGATCTTCAACGCTTTGGTAGATACATTCACCAGAAGAACCATTGACATAAGAC
CTTGAACACCTTTATCAATGAAGTATTGA

ORF Predictions:

ORF #	Start	End	Direction	Length
1	169	678	R	170 aa

>[SEQ ID NO:97] 3174972-1 ORF translation from 169-678, direction R
VIYNIPQLAGVALTPSLYTEMLKNPRVIGVKNSSMPVQDIQTFVSLGGEDHIVFNGPDEQ
FLGGRLMGARAGIGGTYGAMPELFLKLNQLIADKDLETARELQYAINAIIGKL TSAHGNM
YGVIVKEVLKINEGLNIGSVRSPLTPVTEEDRPVVEAAAALIRETKERFL*

Description:

N-ACETYLNEURAMINATE LYASE SUBUNIT (EC 4.1.3.3) (N-ACETYLNEURAMINIC
ACID ALDOLAS E) (N-ACETYLNEURAMINATE PYRUVATE LYASE) (NALASE). -
ESCHERICHIA COLI.

Assembly ID: 3175138

Assembly Length: 1450bp

>[SEQ ID NO:11] 3175138 Strep Assembly -- Assembly id#3175138
CTCCATATTTCTTAGCCTTCTCAATTAGGGTCTTGAAGTCTTCGACACCACCGATACGCT
TACCAATATCAGCATAGTTCAAGTGACCAGAGTCATGGCTGTGATATCCTTAACCTTTTC
CCAACCTTGAGGGTTGTTTCAATGCTACGATAAGCAATGGCACCATCTTGCCAATCAAC
TTTCTTGTCTGCATTGGCATCTTCAGTGATAACAACCTTAGCACTTGGAAGTTCCTTCGT
GTATTCTGGGAAAACAATGCCCTTATAAGCTTTTTTCCCATTGCCATTCAGAGCTGTGGAT
TCCTACATAGTTGGCATTTCGACTGTTTCTTTATAAGCTGTCAAACGAGTCCAGTCATT
CGAACCACCACCATAGCTATTTTGAGAGTTACTCCAAACACCAGCAGCAAGCTTATCTGT
AGAAACAAATCCATACATGTAACCCTTAGCCAAATCCTTCATTGGATTGGTTACATCGAT
ATGATCATCTCCGCTGACATGCGTATTGTTTGACATGGTTGCCCCATCAAACCTTAGCACC
AGTTTGATCACTAGAAACAGAGACTAAAGCATTGCCGAGGAAACTAATAGAAGAAAGTAG
TTTTCTTTTCGTCATCAATCTTTTGACCTGGAGTGACTTGATTGTGGTTGACAATCTTGGT
CACATCAAAGTGCAATTGATTGTCCACAACCTTGCAAGCGTACTGTCATTTCCGCATTGAT
TAAGTGAGCATCATCGCGAAGCTTCATCAAGTACTCTGCTGTTGTCTCATTTGATTTTTTT
ATAAGTGACTTCAGGGGTGATTCGGTGGTTATTGATAAAGACTTGGGTTGAATTGTTGCAC
CTGTCTTGCGAAAGTATGTCCATTCAAGGTGTATCCCTTGACACGAAGGAAGGCTTGGTCT
AATTACTGCCTTAAGTACCTTAAACTGGATCGTATCATAAGTCACCTTGCTATCGTCAAC
AACCGGACCTGTTTCTTTCTGGGCAGGGGTATCCTCTGGGTTTACCCTCTCTGTGGCTA
TCCGTTTCAACGCTTGAACAACCTGGTCGCTCATCGTCATAAGAGCCCGCCTTGAGAAAAA
TCTTCTTCTCATTTCTAAGATGGTCATTGACCGCAGCTGGTAGAGTCACTGTGTCAAAGA

AGATTGACATCCTTATTTGCCTGGCATTACCTGACCGTCTGACTTGAAGACTGATAGAG
 AGACGGTTTGTGATCCTGTTTCAGGAGCAGCAACACGACTACCTCTATACCAAGTGCTA
 GTTGTGGAGATTTATACTCCCAGAACCCAGCCATCCTTGTCATAACCGACAAAAACATTA
 TTATTGGTATCTTTAAATTTCAAGGAGACACCAAAGCGTGATTTGCCCTTTTCAGAATCT
 TCTTTGAAGGTTAAATCAACAGTTGCATTTCCATTGGCATCAACGGTCAAGCCCTTCTTT
 TCAAACAGAG

ORF Predictions:

ORF #	Start	End	Direction	Length
1	79	945	R	289 aa

>[SEQ ID NO:98] 3175138-1 ORF translation from 79-945, direction R
 VTYDTIQFKVLKAVIDQAFRLVKGYTLNGHTLPGQVQQFNQVFINNHRITPEVTYKKINE
 TTAEYLMKLRDDAHLINAEMTVRLQVVDNQLHFDVTKIVNHNQVTPGQKIDDERKLLSSI
 SFLGNALVSVSSDQTGAKFDGATMSNNTHVSGDDHIDVTNPMKDLAKGYMYGFVSTDKLA
 AGVWSNSQNSYGGGSNDWTRLTAYKETVGNANYVGIHSSEWQWEKAYKGIVFPEYTKELP
 SAKVVITEDANADKKVDWQDGAIAYRSIMNPNQGWKVKDITAMTLVT*

Description:

unknown

Assembly ID: 3175860

Assembly Length: 420bp

>[SEQ ID NO:12] 3175860 Strep Assembly -- Assembly id#3175860
 CTGCGAGTTGTGAGGCTCCTATTATGTCTCGTGATTAAAATCTCTATAAGGTGATTTTGG
 AGGGAAATTATCGGGCGACAGCGGGTAGAGAAGAGATGAAAGAGGCTATTTTGGAAATATC
 AAGCAAATCCTGCTGCCTTAAAAGATCTCAAAGAAAAGGCTAAGAATATTTCCAGAGAGT
 ATTCTGAAGAGCATCTGTTACAAATCTGGTTGGACTTTTATGAGAAACAAGCCGCTTTAG
 GGACAAAGTAAAAAGTGAGGTAATCTATGCGAATTGGTTTATTTACAGATACCTATTTTC
 CTCAGGTTTCTGGTGTTGCGACCAATATCCCAACCTTGAAAACCCACCTTGAAAACACGG
 ACTTGCCCTGCATTTNTATCTCATACAATCCACCGAATTTTCGATGTCCCCCTCCCTACAAC

ORF Predictions:

ORF #	Start	End	Direction	Length
1	51	251	F	67 aa

>[SEQ ID NO:99] 3175860-1 ORF translation from 51-251, direction F
 VILEGNYRATAGREEMKEAILEYQANPAALKDLKEKAKNISREYSEEHLLOIWLDFYEKQ
 AALGTK*

Description:

unknown

Assembly ID: 3175918

Assembly Length: 661bp

```
>[SEQ ID NO:13] 3175918 Strep Assembly -- Assembly id#3175918
CTCCCCAAACTTTTATTTGAGAGTGAACGGTATAAGAATATGAAACCGGAGGTTAAGGTG
GTTTACTCAGTTTTAAAGATCGGTTGGAGTTGTCTTTGAGCAAAGGTTGGATTGATGAG
GATGGGACTATTTATTTGATTTATTCCAATTCAAATTTGATGGCACTTTTAGGCTGTTCA
AAGTCAAAATTACTCTCCATGTGAGTTTGAAGTGACATTTTATAGATGATTACCATAAAAA
ACATAACTACCCACTATTTTACGAATCCTATCTTCAAACGTTATGGAATTCCTTGAAAG
TCAAGACATAAAGAATGGGGTTGATGCCTTTGTAGATGATCATCAAATCTCGTTTTTGT
TTTATATGGACAAGGCTATCGAGCCGAGGGAAAAGAGGGAATACTTACAACCCAAGTAAC
TGTAAGCTTATGATGAAGACAAGAAACCGATTAACCTTCGCAAATTTATTAGATTCCTT
AATCGTGTCAGAATATCAAATGGAACCGAATCTTTGGGAGGTCTCCTATGATTGATCTCT
ATCTAAGTAAAAATAGCCGAAGAAATCAACTTCTTTTAGACTTCTTCCAAACTATGGCA
TCGAGGTATCTTGTTCATTTCAGTTTCTGAAATGACAAAGGACAAATTAATTGAGATGATGA
G
```

ORF Predictions:

ORF #	Start	End	Direction	Length
1	212	535	F	108 aa

```
>[SEQ ID NO:100] 3175918-1 ORF translation from 212-535, direction F
VTFLDDYHKKHNYPLFYESYLQNVMEFLESQDIKNGVDAFVDDHQNLVFLVLYGQGYRAEG
KEGILTTQVTVKAYDEDKKPINFANLLDSLIVSEYQMEPNLWEVSYD*
```

Description:

unknown

Assembly ID: 3811220

Assembly Length: 1429bp

```
>[SEQ ID NO:14] 3811220 Strep Assembly -- Assembly id#3811220
CTGCCCCCTGTAAGGCTGGACGATTGCCTTTCTTAGTATCCGCAAAGAGGTAACTGAGAA
TAGAGAGGATTTCTCCTTCAATATCTTTGACAGACAGGTTTCATCTTGCCCTTCTACGTCTG
AAAAAATCCGCATATTGACCAGTTTTTCTCACAGCATAGTCCAAATCTTCCTCTTGGTCCT
CTGGTCCAACACCAACCAGCAATAAAAGTCCCTGATTGATTTTCCCTGAATCTGGCCTT
CTATACTCACTTGGGCTTTTTTAACCCGTTGGATAATGATTTTCATAATAGCCTTTCTAG
TAAGAGCTAGGACAACTAGCCGTTGGTCCGTTTGACAGAGTAACTTCTGGCACACTCTT
AATTTTATCGACAACCGTGGTCAGTGTAGAGAGGTTGGCAATACCGAAGGACACATGGAT
```

ATTAGCAAAC TTCATATCCTTGGTTGGTTGGGCATTGACCGTTGAAATATTCTTGGTTGT
 ATTTGAAAGAACTTGCAGTACATCGTTCAACAGTCCTGTACGGTTGAGACCGTAGATATC
 GATATGGGCCATATACTCCTTATTTGAGCTAGAGTACTGGTCTTCCCATTCCACATCAAG
 GAGACGTTGCTCGTAGTTTTCTTGGGCACGCAGGTTTCATACAGTCCACACGGTGAATAGC
 CACACCACGACCCCTTGGTAATGTAGCCAACAATATCGTCAACCAGGCACGGGGTTACAACA
 CTTAGCAATCCGCACTAGGAGACCAGAAGCACCTTCAATAACCACTCCCCCTCATGCTT
 GACCTTGGAGAGTTTCTTTATTTTCAACCTTGACCTCGCCACCTTTGACAAGCTCCTCTG
 CCTCAGCCTTGGCCTTGGCAGCTCTTCTCACGGCGTTCTTTTTTCAGTCAGACGGTTAA
 AGACGGTAATCGCACCATTTCCTTCCCAAACCAATGGCCGCAAAGAGGGAGTCTTCTGTCT
 TGTAAC TGGTCTTTTGCAGAACTTGATCCATGTGGCGCTTGTCATAAAATTTATTTGCCA
 CATAGCCATTTTCTTGGAAGTGAAGCATCAGCATCTCACGACCCTTGTTGACAGACAATT
 CCTTATCTTGGTTTTTAAAGAACTGGCGAATCTTATTGCGCGCCTTGCTAGTCTTGACCA
 TATTGAGCCAGTCACGGCTAGGTCCAAAGGAGTTCGGGTTGGCGATAATTTCAACCTGAT
 CCCCTGTCTTTAACTTGGTTGTGAGTGAACCATGCGGCCATTGACCTTGGCACCAGTTG
 CTTTTTCACCGACCTTGGTATGGATTTTCGTAGGCAAAATCAATCGGTCTGAATCTTTGG
 GAAGAGAACGGACAGCTCCATCTGGGGTAAAAACGTAAATCTCCTCAGCCAGATAGTTTT
 CCTTAACAGAGTCCACAAATTCCTTAGCATCATCAGCCTGGTCTTGGAG

ORF Predictions:

ORF #	Start	End	Direction	Length
1	316	873	R	186 aa

>[SEQ ID NO:101] 3811220-2 ORF translation from 316-873, direction R
 VRKSVPRRLRQRLSKVARSLKIKKLSKVKEGGVIEGASGLLVRIAKCCNPVPGDD
 IVGYITKGRGVAIHRVDCMNLRAQENYEQRLLDVEWEDQYSSSNKEYMAHIDIYGLNRTG
 LLNDVLQVLSNTTKNISTVNAQPTKDMKFANIHVSGIANLSTLTTVVDKIKSVPEVYSV
 KRTNG*

Description:

stringent response-like protein - *Streptococcus equisimilis*

Assembly ID: 3811436

Assembly Length: 1513bp

>[SEQ ID NO:15] 3811436 Strep Assembly -- Assembly id#3811436
 CTCTGCAATGATGTACTCAAACATCTCCGCTTCTAGTTCCTCCTTAGGCAGAGGCAATTT
 CCCACGTCGCATCCGGTTCATAAAGACCGTATGGTTTTCTAAATCAAAC TATACAAAC
 CATGTGGGGAATATCCAATCCAATGGCTTTAGCCACATTTTCTTTACTTGCTCCATGGT
 CTGACCAGGCAGAGCATAAATCAAATCAATGGAGATGTTGTCAAACCAGCCAGTTTCAG
 GCGATCGATATTTTCATAAATATCCTTCTCAAATGACTGCGCCAATCTTTTCAACAT
 CTTATCATCAAAGGTCTGGACACCTAGCGAAACACGATTGACAGCCGAATTTTCAAAC
 AGCTATCTTATCCGCATCCAATCGCCTGGATTGGCTTCAATGGTCAACTCTTCCAAGAC

AGACAAATCCAAGTTTTTTAGTCAAGCCATTTCAGTAACACCTCCAGTTGCGGAGCCGACAG
 GGCTGTCGGTGTTCACCACCGATATAAAGGGTTGACAACTTTTCAATATCATAAGAACG
 AAACCTCTTCCAGCAGATGCTCTAAATAGCTGTGCGACTGGCTGATTTTTTGATGAAGACCTT
 TGAAAAATCACAATAATAACAAATCTGGGTACAAAATGGGATGTGCACATAGGCTGACGT
 TGGTTTTTTCTGCATAGTAATTATTATACCACAAAGACTAGATTCCAGATAAAAAATCACC
 ATCCCCAGATACATAGTCCGTCCGGAGATGGTGATGGTTTATTCTTCTGTTATATCAATC
 ACAATCTCTTCTGAGTCATCAAGAGCTTCGGCTTTTTCTTGCCATTGTTTCCTTGAGATTA
 TTTAATTGATTTTTTTGATGCTTCTGTGCGCTTGAAAAGCATAGGATTTAGCTTGAGCAAGT
 ATACTGTCCACAGTGATTTTACCTGACTCAACCTGTTCTTTTGTTTTCAGAACAAAATCT
 GTAGCCTGCTCCTTAACCTTCTGTGAGTTTTTTCACAGACTTGCTCCTTGGCATACTCCGGA
 TCTTCTCTCAAATCATCTAAAAATCTTGAGCCTGACTGCAAACTTGTTTGCCCTTATCA
 CTTGTTAAAAACAAGGCAAGAGCTGCACCTGAAACGGTTCCTAAAAGGATTGAGGATAAT
 TTACCCATAAGGATTCTCCTTTTTTTATTTTTTGAAAAATTTACTTGCAAGACGAAGAGCT
 GACAGACTTGACACCAGTCTTGAGTGTTTTTGAACCAGCTGATGAAGCTTCTTGCTCAAG
 ACACGCGCATGGTCATTGAGGTCTGAAACAGATAGAGATAAATCTGCAACAGCACTGAAG
 AGTGGATCAATCGTAGCCACCTTGACATTGATATCATCTGCCAAGACATTGACCTTAGCC
 AACAACTCATTGGTGTGATGCAAGGTCACATCCACATCTGAAGTCAAGGTTTTAATCGTC
 TTTTCTGTTTCATCGATGACACGACCAAGCTTTTGTACAGTAATGATCAGATAGACCAA
 AAGACAATCACAG

ORF Predictions:

ORF #	Start	End	Direction	Length
1	1164	1511	R	116 aa

>[SEQ ID NO:102] 3811436-3 ORF translation from 1164-1511, direction R
 VIVFLVYLIITVQKLGRVIDETEKTIKLTSDVDVTLHHTNELLAKVNVVLADDINVKVAT
 IDPLFSAVADLSLSVSDLNHDHARVLSKKASSAGSKTLKTGASLSALRLASKFFKK*

Description:

unknown

Assembly ID: 3811984

Assembly Length: 505bp

>[SEQ ID NO:16] 3811984 Strep Assembly -- Assembly id#3811984
 CTCTTGTCAGAGAAATTTACAAAACGTTAGGAGAATAAGATGGCATTTATTGAAAAAGGT
 CAAGAAATCGATATGGAAGTCATCAAGGCTGAAACCCAATTGTCTGCAGAAGCCTTGAGA
 CTCAAGGAAAGCCGTGACAGGGAATTGGCAGATATTATTTTCAGGGGAAGATGACCGTATT
 CTCTTGGCTGATTGGTCTTGCTCTTCTGATAATGAAGAGGCGGTCTTGGAATATGCTCG
 CCGTTTTATCCGCCTTGCAAAAGAAGGTAGCGGATAAGATTTTCATGGTCATGCGCGTGTA
 TACTGCTAAGCCTCGTACCAATGGAGACGGCTATAAAGGGTTGGTTCACCAGCCAGATAC
 TTCTAAGGCTCCAACCTGATTAAACGGCTTGACAGGCTGTGCGCCAGTTGCACTACCGCGT

TGATTACAGAGACTGGTTTGACAACGGCAGATGAGATGCTTTATCCGTCAAATCTGATCT
TGGTGGATGACTTTGGTCACCTACC

ORF Predictions:

ORF #	Start	End	Direction	Length
1	134	454	F	107 aa

>[SEQ ID NO:103] 3811984-2 ORF translation from 134-454, direction F
VTGNWQILFQGKMTVFSWLIGPCSSDNEEAVLEYARRLSALQKKVADKIFMVMRVYTAKP
RTNGDGYKGLVHQPDTSKAPTLLINGLQAVRQLHYRVDYRDWFDNGR*

Description:

PHOSPHO-2-DEHYDRO-3-DEOXYHEPTONATE ALDOLASE, TYR-SENSITIVE (EC
4.1.2.15) (PHOSP HO-2-KETO-3-DEOXYHEPTONATE ALDOLASE) (DAHP
SYNTHETASE) (3-DEOXY-D-ARABINO-HEP TULOSONATE 7-PHOSPHATE SYNTHASE).
- ESCHERICHIA COLI.

Assembly ID: 3857228

Assembly Length: 1827bp

>[SEQ ID NO:17] 3857228 Strep Assembly -- Assembly id#3857228
CTCTTTTAAACCGTTTTAGCGGTGACACCGAGGATTTTTTCAGGACCCAAGACTTGTCGGG
CAACCGAAACTGGGAGTTCGTCATCTCCAATATGCAGACCAGCAGCATCAACCGCAAGAC
AAACATCCAACCGATCATCGATTATCAAGGGGACCTGATAGGCATCTGTTATTTCCCTTGA
CTTGTTTTTGCCAGTTGATAATATTGATTGGTTGTGAGATTTTTTCTCGCAATTGGACTA
TGGTAACCCCTGAACGGCAGGCCGTCTCAACTTTTGCAAGAAAGCTTCCACGGAATCTT
GATAGCGATTGGTTACCAGATATAGTCTAAGCGCTTCTCTATTCATAAACCTCTCCTTTG
ATGGTATCTAGCCAATTTTCATCTCTTCTTAGGAGCGAAAGCTGATTGAGTACTTGGTAA
CGAAATCTTCCAATCCCATTCCTTGAACAACCTATTTTCTCAGCAGCGATATTGAGATAA
GAGACTGCTAAGCAAGAACTTCAAAACCAGTCTTTCCTTGGCTGAGAAAAACAGCTGTTA
AGGCTCCAACCAAGTCTCCTGTCCCTGTTATCCAGTCTAATTCAGTACAGCCATTCTCAA
GTACAGCAACTTGATTCTCCGAAACAATAAGGTCCCTGGGACCTGTGACTAAGAATGACA
TACCACGATAGGTCTGACACCAGTCTTTCAAGACTTGAAGCAAATCCTCCGTTTCTTGAT
CTTTAGCACTCGCATCGACCCCAACGCCGTGATGCTTTAATCCAACAAGACTTCGAATTT
CTGACATGTTTCTTTAAGGACCGTAGGTCTATAGTCTAAAAGGTCTTTAACTAAGCTCT
TACGAATGGATGAAGTCGTTACGCCAACCGCATCTACTACCATCGGGAGAGAAGATTGGT
TTGCATACAAAGCTGCCATGCGGATTGCTTTTTCTTCTCAGCTGACAAATGCCCCAAAT
TGATGAAGAGAGCCTGGCTTTGCTTAGTAAAATCAAGAACTTCACGGGGATCATCTGCCA
TGACAGGTTTGCATCCCAGAGCCAAAATCCCATTTGCCAGCATCTCACAAGAAATCTCAT
TGGTCATACAGTGAATGAGGGAAC TAGAGCCTATAGGAAAAGGATTTGTCAATGCCTGCA
TCATTCTATCCTTTTCAGCAAAGAAATATCCTTGCACTTTTTTAAAGAATTCTTGCTTGAT
TAAAAATCTAAATGCAATAAAGGAAATCGCTGTACCAATCAAGGTTGCTCCGAAAAATCG

AGGCGTGTAGATAAACCAACTAAGCTTAGCAGCCGATCCTGTAAAGAGCACCATAACAGG
 ATAGGAAACAATAGAACCAATAATACCTGTTCCCACAATTTCTCCCAAGGCAGAAAAGTA
 AAATTTTCGACCGTACTTATAAAAGAGACCTGCTAGAAGGGCTCCAAAAGTCGCTCCTGT
 GAGAGATAAAGGAGCTTATCGGAATACCCCTTGAGTCGTCATACGGATAAAGGCTGTCACT
 GTAGCCATAGCCAAGGCATAAACAGGTCCCATCATGATTCCCGCTAGAATATTGACTACA
 CTGGACATCGGTGCCATTCCCTCAATCCGAAAGATAGGTGTAAGGACTACATCAAGGGCA
 ATCATCATAGATAAAATGGTCAATTTGTGAACTTGTAGTTGGTGCTTTCTCAAGTTTCTA
 TTCTTCTCCTTTTTTCTAAAGACTGTAAATCGCTCTTCCATGTCTGGTGTTGGTAAGCCAT
 CTCCCAAACTTGGCTTCCATATGAACACTGATGTGGAAGGCATCTAGCATTTTTTTGCTT
 ATCTGTCTCATCACTTTCTCGATAGAG

ORF Predictions:

ORF #	Start	End	Direction	Length
1	1141	1356	R	72 aa

>[SEQ ID NO:104] 3857228-2 ORF translation from 1141-1356, direction R
 VGTGIIGSIVSYPMVLFTGSAAKLSWFIYTPRFFGATLIGTAISFIAFRFLIKQEFFKK
 VQGYFFAERIE*

Description:

unknown

Assembly ID: 3857842

Assembly Length: 485bp

>[SEQ ID NO:18] 3857842 Strep Assembly -- Assembly id#3857842
 CTATTGCCAATCCATATAGCCTATCAGGTGGTCAATAACAACGTGTGGCCATCGCTCGTG
 GCCTATCAATGAATCCAGACATCATGCTCTTCGATGAACCAAATTCTGCCCTTGACCCTG
 AGATGGTTGGAGAAGTAATTAACGTTATGAAGGAATTGGCTGAGCAAGGCATGACCATGA
 TTATCGTAACCCATGAGATGGGATTTGCCCGCCAGGTTGCCAACCGCGTTATCTTTACTG
 CAGATGGCGAGTTCCTTGAAGACGGAACACCTGACCAAATCTTTGATAACCCACAACACC
 CTCGTCTGAAAGAGTTCTTAGATAAGGTCTTAAACGCTTAAACTCAAACCTGCAAGGATTT
 CCTTGACGTTTTTCTACCTCGTATTGGAATTTTTTGATTTTTTCGGAAAATTATGTTAGAAAT
 TAAGTTTATGAAATGAGGTTTCCTCATACCTAGCAAGACTAGGAATAAAAATAGAAATTA
 GG TAG

ORF Predictions:

ORF #	Start	End	Direction	Length
1	45	341	F	99 aa

>[SEQ ID NO:105] 3857842-1 ORF translation from 45-341, direction F

VAIARGLSMNPDIMLFDEPNALSALDPEMVGEVINVMKELAEQGMTMIIVTHEMGFARQVAN
RVIFTADGEFLEDGTPDQIFDNPQHPRLKEFLDKVLNV*

Description:

GLUTAMINE TRANSPORT ATP-BINDING PROTEIN GLNQ. - BACILLUS
STEAROTHERMOPHILUS.

Assembly ID: 3857996

Assembly Length: 1547bp

>[SEQ ID NO:19] 3857996 Strep Assembly -- Assembly id#3857996
NTCTTGGGCNCNCGGCGNNTCCTTTGAGGACNACGGTATCGATGACCTTGATCTCAAGTG
CAAGCAGTATCTGAATCTGCAGCAGCACCTGTCCGTGCAAAAGTTCGTCCAACATACAGT
ACAAACGCTTCAAGTTATCCAATTGGAGAATGTACATGGGGAGTAAAAACATTGGCACCT
TGGGCTGGAGACTACTGGGGTAATGGAGCACAGTGGGCTACAAGTGCAGCAGCAGCAGGT
TTCCGTACAGGTTCAACACCTCAAGTTGGAGCAATTGCATGTTGGAATGATGGTGGATAT
GGTCACGTAGCGGTTGTTACAGCTGTTGAATCAACAACACGTATCCAAGTATCAGAATCA
AATTATGCAGGTAATCGTACAATTGGAAATCACCGTGGATGGTTCAATCCAACAACAAC
TCTGAAGGTTTTGTTACATATATTTATGCAGATTAATTTACAGAGGGACTCGAATAGAGC
CCTCTTTTCAGGTTTACCGTGACAATCCCTATTAAAAATTATATCAAATCGTAAAAAT
ATTGGAAGATATGGTAGAATGAAAATTGTCGTGTGAACGATAATACTCATTCTTGATGA
ATTGTGAAGCAGTTGCCCTTGGGTCGTTTTCGAGTTGAAGTCAAGAAGAGGAAAAAAC
AAAAAGGAGAAATACTCATCGAATTTCAATGAAACAACCTTCTTGAGGCTGGTGTACACTT
TGGTCACCAAACCTCGTCGCTGGAATCCTAAGATGGCTAAGTACATCTTTACTGAACGTAA
CGGAATCCACGTTATCGACTTGCAACAACTGTAAATACGCTGACCAAGCATACGACTT
CATGCGTGATGCAGCAGCTAACGATGCAGTTGTATTGTTTCGTTGGTACTAAGAAACAAGC
AGCTGATGCAGTTGCTGAAGAAGCAGTACGTTTCAGGTCAATACTTCATCAACCACCGTTG
GTTGGGTGGAACCTTTACAACTGGGGAACAATCCAAAAACGTATCGCTCGTTTGAAAGA
AATTAAACGTATGGAAGAAGATGGAACCTTTCGAAGTTCTTCCTAAGAAAGAAGTTGCACT
TCTTAACAAACAACGTGCGCGTCTTGAAAAATTCTTGGGCGGTATCGAAGATATGCCTCG
TATCCCAGATGTGATGTACGTAGTTGACCCACATAAAGAGCAAATCGCTGTTAAAGAAGC
TAAAAAATTGGGAATCCCAGTTGTAGCGATGGTTGACACCAATACTGATCCAGATGATAT
CGATGTAATCATCCCAGCTAACGATGACGCTATCCGTGCTGTTAAATTGATCACAGCTAA
ATTGGCTGACGCTATTATCGAAGGACGTCAAGGTGAGGATGCAGTAGCAGTTGAAGCAGA
ATTTGCAGCTCCAGAACTCAAGCAGATTCAATTGAAGAAATCGTTGAAGTTGTAGAAGG
TGACAACGCTTAATTTATACAAATAGTAATTACCTAGGAGGGCGGGGCTTAGCCCGGCTC
TCCTATTTTCAAAAAATATAGGAGAATTAAATGGCAGAAATTACAG

ORF Predictions:

ORF #	Start	End	Direction	Length
1	58	456	F	133 aa

>[SEQ ID NO:106] 3857996-1 ORF translation from 58-456, direction F
 VQAVSESAAAPVRAKVRPTYSTNASSYPIGECTWGVKTLAPWAGDYWNGAQWATSAAAA
 GFRTGSTPQVGAIACWNDGGYGHVAVVTAVESTTRIQVSESNYAGNRTIGNHRGWFNPTT
 TSEGFVITYIYAD*

Description:

unknown

Assembly ID: 3858236

Assembly Length: 740bp

>[SEQ ID NO:20] 3858236 Strep Assembly -- Assembly id#3858236
 CTATAAAAAAAGGGTAACCAGTATGGAGGATGAATGTCTGGAACCTATCTGAGAATCTCG
 GATTTTGGAAATCAGACCGATCATCATGAGATAAGGAAGGAAAGCACTTGTA AAAAGCAC
 TGTAACCACGCCAGTCCCCTGTCCCAAGAGGGTGAGGTGGTAGCGTAAAACCATGCGGAA
 AAATCCCTTTTTAGTGGTTGAAATTCTCTCCTTGCTGCGACGTTCTTTTTTGACCTTCTC
 CTCCTATTAAGCAGGATCACGTCATAAAAACGAGGAAGGACCTTCTTTTTTGCTCAGATA
 AAGCAGGAAGAGAGTTAGTCCCTATCCAAGCGAGCAGACCCAATATGGCTTCTATTGAAAA
 AGGCTCCACTGCTATTTTGTAAAAGATATGAAGAGGATAAAGGAGAAATGGAATGTCTCT
 AACTTTGTCAACAATACTTCCAAAAGTCGACTGAAGAAAGAAGATAAATATTAAGGTAT
 GAGAACTCCTATCCCAATCATCACATTGCAAAAAATAGACTGATACTTTCTGAAGACCCT
 AGTCTGAGCCAAGAAATGTACTGCCACTACCGTCACTAAAGTAACAGAGACAAATAATAA
 GGTCAAGGACAGTAGCATCAAAGGCAAACCCAGCCAAAGAGAAGGAGCTAGACTAATATA
 GAGGGCTAGAAAATAAGCTAGGATTGGTACAATTCCAGTTAGAGCTGGCAAGAGGACAGA
 CAGTCCTTTAGCAATTCGAT

ORF Predictions:

ORF #	Start	End	Direction	Length
1	1	261	R	87 aa

>[SEQ ID NO:107] 3858236-1 ORF translation from 1-261, direction R
 VILLNSEEKVKKERRSKERISTTKKGFFRMVLRHYHLTLLGQGTGVVTVLFTSAFLPYLMM
 IGLISKIRDSQIVPDIHPPYWLPFFL*

Description:

unknown

Assembly ID: 3858264

Assembly Length: 2219bp

>[SEQ ID NO:21] 3858264 Strep Assembly -- Assembly id#3858264
 ATCGAATTCGTTTTGCAAGTGGCGAAATGCGAACCACGTTTGTGTCTTTATAAGTTTCCA

CGTCTTCTTTGTGGACACGACCGTTTGCACCTGAGCCAGAAACGTCGTAGAGGTTTATCC
 CTAAATCATCCGCTAACTTTCTAGCTGCAGGAGTCGCTCTTAGCTTGTCATCAGCCATGA
 CCTCTCCAATTCTATTTATGATACAAAGGGCGTCAAAAGCGACTGAAAAATAGGAAATCG
 ACGATGGCTTCGATGAAGCCAAGGAGATTTATCTTTTTTTTCCAAGCTTTTAGCCCGTGCT
 CTAATCTAAGATATTAAGGACGAAGAGCTCTGCACCTAAAAGATACAAAGTTCTCGTCAG
 CTTTGTTTTATTTACATAACTTATCTTATGTAACCTCTATTCTTTGTTATAAGTTTTTCGG
 ATTGCATCTTTGATACTTTCAACTGTTGGAATCATTGCACATTTTTTAGGTTTTGCGCATA
 AGGCATCGGCACATCTTCTCCTGCACAACGGCGGATTGGTGCATCTAGATAGTCAAATGC
 TTCTGATTCTGAAATAATAGCTGAAATTTACCGATATAGCCACTTGTTTTGTGGGCATC
 GTTGACCAGAACAACCTTACCAGTCTTCTTCACTGAGTTTATGATGATATCCTTATCAAG
 CGGAACAAGGGTACGTGGGTCAACAATTTCAACTGAAATTCCTTCTTCAGCTAATTCCTC
 AGCAGCTTGAACCACACGGCGAAGCATTTTTCCATAAGTGACAACCTGTTACATCCGTTCC
 TTGGCGTTTTGATTTACCAACCCCAAGTGGAATTGTGTAGTCTGGATCAACTGGCACTTC
 CCCTTTTTTGGTTAAATTCTGACTTGTACTCAAGTATAATAACTGGGTGTTATCACGGAT
 AGAAGACTTAAGCAGGCCTTTTCATGTCCGCAGGTGTTCCAGGTGCCACAACCTTAAGCCC
 TGGAATGTGAGTAAACCAAGACTCTAGAGATTGTGAGTGCTGGGCGGCAGAGCCAACTCC
 GTTACCAGCTGCACAACGAACAGTCATTGGAACCTGACCTTTACCACCAAACATGTAACG
 TGTTTTAGCAGCTTGGTTGACGATATTGTCCATGGCAATAACAGAGAAGTCCATGAAGGT
 CATATCGACGATTGGACGAAGTCCTGTCATGGCTGCTCCTGCTGCAGCTCCAGAGATGGC
 AGCTTCAGAAATCGGACAGTCACGGACACGTTCTGGACCAAATTCCTCAAGCATTCCAAC
 AGAAGTACCGAAGTCTCCTCCGAAGACACCGACGTCCTTCTCCCATCAAGAACACATTTTC
 ATCGCGAACGCATTTTCCTCAGACATAGCAAGGATAATGGTGTACGGAAGGACATTGTTT
 TTGTTTTCCATTTTATCTCTTTCTCCTTAGTCTGCGTAAATATCTTCAAAGGCTGATTCAA
 GCGGTGGGAATGGGCTTTCCTCTGCAAATTTAACAGAAGCTTCTACTGCTTCCTTTACTT
 GCGCTTGGAATTCCTCCAATTCTTCGGCACTTGCAATGTTATTTTCAATAAGGTAATTGC
 GGAGGTTTTTCGATTGGATCTTTTTGTTTCCACAATTCCACTTCTTCACGCGTACGATATT
 TACCAGGGTCAGATGATGAGTGACCGAGCCAGCGATAAGTTACACTTTCAATCAAGACTG
 GACCATTGCCACTGCGAACATGGTCTATAGCTTTCTGAAATCCTTCATAGACATCGATGA
 CATTGTTACCGTCTTCGATGAACATTCAGGAATTCATAAGCGGCGTACGTTGATGGA
 TATGTTCTATATTGGTCATTTTCTTGATATCCGCAGAGATACCGTAACCGTTGTTAATGC
 AATAGAAAATGACTGGCAGGTTCCAGATAGAAGCCATGTTCACTGCTTCGTGGAAAACAC
 CTTCAATTGGTCGCACCATCTCCAAAGAAGCAGACAACGATTTTACCGGTATTTTGCATTT
 GCTGACTGAGGGCTGCACCGACAGCGATCCCCATACCACCACCTACGATACCATTTGGCAC
 CAAGGTTCCAGCATCAAGGTCAGCGATATGCATAGATCCACCTTTCCCTTTACAGGTTTC
 CAGTGTATTTACCAAGGATTTTCAGCCATCATTCGTTGAAGTCAATCCCTTTAGCAATAG
 CTTGCCCCGTGTCCACGGTGGTTTGAGGTAATCAGATCATCTGGATTGAGAGCTACATAG

ORF Predictions:

ORF #	Start	End	Direction	Length
1	439	1365	R	309 aa

>[SEQ ID NO:108] 3858264-1 ORF translation from 439-1365, direction R

VTPLSLLCLRKCVRDENVFLMGEDVGVFGGDFGTSVGMLEEFGERVRDCPISEAAISGA
 AAGAAMTGLRPIVDMTFMDFSVIAMDNIVNQAAKTRYMFGGKGQVPMTVRCAAGNGVGSA
 AQHSQSLESWFTHIPGLKVVPAGTPADMKGLLKSSIRDNNPVIILEYKSEFNQKGEVPVD
 PDYTIPLGVGEIKRQGTDTVTVTYGKMLRRVVQAAEELAEEGISVEIVDPRTLVPFLDKDI
 IINSVKKTGKVVLVNDAAHKTSGYIGEISAIISESEAFDYLDAPIRRCAGEDVPMPYAQNL
 KMCNDSNS*

Description:

2-OXOISOVALERATE DEHYDROGENASE BETA SUBUNIT (EC 1.2.4.4) (BRANCHED-
 CHAIN ALPHA -KETO ACID DEHYDROGENASE COMPONENT BETA CHAIN (E1))
 (BCKDH E1-BETA). - BACILLUS SUBTILIS.

Assembly ID: 3858610

Assembly Length: 1078bp

>[SEQ ID NO:22] 3858610 Strep Assembly -- Assembly id#3858610
 CTAACCCCTNGACGGGGCCGCTATCATCAGTCAAACAGCTAAAAATCTTGCTGCAAAAGT
 CTCGATTAACTGAGCTTTTACAAAAGCCGTATTTCTGGAATAACTTGGAGATTGATCAT
 CTTATCCATCAATTCAGCCGATTCGATATTGTCTTCAGCCAGTTGCAGACTTTTTACGAT
 TGATTTTGGCAATTCGTAGACATAGGTGTTGTCTCTCAAAGGAATTTTGACAATACCTAA
 CTCTTTGATATCTCGGGATACCGTCGCCTGAGTGGCAGTGATACCTGCTTCTTTCAAATG
 TTCTACAATTTCTTCTTGCGTGCCGATTTGATAATCTGTACCAATCTTCTAATTTTTTC
 AAGTCTCTCTTTTTTATTTCATTTTTTAAATTGACTATGCGCCCTCTCTACTGCTTCTTTAA
 TCTCAGCAAGAATCTGATTGCTTGCTGACTTTTCTTTTTTCAAATACACTAAAAATTCAA
 TATTTCCATGTCCACCTTGGATGGGAGAAAAGTCCAAGCCAAGGACTGAAAAACCTGCCT
 CTACTGCCATAGCTGTTACAGATTCAAGGACATTCTGATGAATCTTAGCATCTCGAATAA
 TTCCATTTTTTCCCAATCTGCTCACGTCTGCCTCAAACCTGAGGTTTGACAAGTGCTACCA
 CCTGACCTTGATCAGCCAAGACACGGTGCAAGGCTGGCAAAATCAGACTAAGGGAAATGA
 AACTCACATCAATACTGGCAAAGCTCGGCTCCTGCTCGAAATCAGTCTTTTCAGCATAGC
 GGAAATTGAACTGCTCCATGCTGACAACCTCGTGGGTCTTGCGTAATTTCCAAGCCAACCT
 GATTGGTACCAACATCGACTGCAAAGACCAACTTGGCACTATTCTGTAGCATGACATCGG
 TAAACCTCCAGTAGAGGCCCGATATCAATCGTAGTCGCGCCATCCACCGACAAATCAA
 AGACCTGCAAGGCCCTTTTCCAGTTTCAAACCACCGGCTGACATACTTGAGTTTCTCC
 CCCTTGAGTTTTTAATTCGGTGTCATCTGGAATTTCTCTCCTGGCTTGTCAAACCGTTC

ORF Predictions:

ORF #	Start	End	Direction	Length
1	374	949	R	192 aa

>[SEQ ID NO:109] 3858610-2 ORF translation from 374-949, direction R
 VDGATTIDIGASTGGFTDVMLQNSAKLVFAVDVGTNQLAWKLRQDPRVVSMEQFNFRYAE
 KTDFEQEPSFASIDVSFISLSLILPALHRVLADQGQVVALVKPQFEAGREQIGKNGIIRD

AKIHQNVLESVTAMAVEAGFSVLGLDFSPIQGGHGNIEFLVYLKKEKSASNQILAEIKEA
VERAHSQFKNE*

Description:

cytotoxin/hemolysin ORF2 tly - *Serpula hyodysenteriae*

Assembly ID: 3858716

Assembly Length: 928bp

>[SEQ ID NO:23] 3858716 Strep Assembly -- Assembly id#3858716
ACTTTCCTGACCTCTGTTTCCAAATAATCTTCCAAATGGACAGAGATCTACCGTTGTTTG
CATCGATAGCTGAGGTCTTTTTTAGAAAATACCATCACTTTTAGAAAATATAAACACATT
TTTCGGATAAGATTAAGGTAAAAGCAGCTCGTTTATCCAGGGTCTGATGATGGTCTTCA
CGATAAACCACATCCAATAACCAATGCATACTTTCTGCTGACCAATGACCTCGAACACTA
TGGCAAAAGGTCATCAACATCAAGCTTAAAGTTAAAGATAAAATAGCGAACGTCTTGACT
TGTAATACCATCTCTATCAATAGTATTACGAGTCATTCCAATTCACGCAATTTATGCCA
TTTGGGATGGTTTTGACACAACCACTTAACATCAGAAGACACCCAGTATTCTCGAACTTC
AATCTATCCTCTTTCTATATTCTAACTGAAAGGACAATTCAATGATTCATTTAATAATGA
TTAGCGCCATTGCTCTAGCCATTGGAATTGGTTACCGCACCAAAATCAATATTGGCCTGC
TGGCTATTGCTTTTTCTTACCTCATCGCAACCACTCTCATGGGATTAAGTCCCAAAGAAC
TTCTTCATTTTTTGCCAACCTCACTCTTTTTTACCATTTTTAGCGTCTCTCTCTTTTATA
ACGTTGCAACAACCTAACGGTACTCTTGATGTTTTGGCTCAACACATTCTCTACCGCACAC
GCACCCACCCTAACGCCCTCTACATGATTTTATACCTGATGGCAACCCTTTTGTCTGCTT
TAGGTGCTGGATTTTTCACTACTATGGCCGTTTGTCTGTCTCTAGCGATTACCCTCTGTC
AAAAAGCGGACAAACACCCTTTGATTGGAGTCAAAGCGTCAATGGGAACTTCAGGAAGGG
TAATTTGATAACCAAAGGAATAAAATTT

ORF Predictions:

ORF #	Start	End	Direction	Length
1	238	402	R	55 aa

>[SEQ ID NO:110] 3858716-1 ORF translation from 238-402, direction R
VSSDVKWLCQNHPKWHKLRGIGMTRNTIDRDGITSQDVRYFIFNFKLDVDDLPL*

Description:

unknown

Assembly ID: 3859124

Assembly Length: 847bp

>[SEQ ID NO:24] 3859124 Strep Assembly -- Assembly id#3859124
AAAAACGCACCATATCAAAAATAAAAGTTTGATATCATGCGTCATGTCTTAACTAAT

TGACTATACTTTCTATTCAAATGAGCTTTTAACCAATTGATTGAGCCAATCCACTCTTAA
 AACCAAAGGAGCAATTTCTCGGCTTAGCTGACTCTTCTCGGAATCTGAACCATGTACAAC
 ATTTTGATAATCTCATTTTCTCCAGCAGCTTTTGCAAAATCACCTCGAATAGTGCCTGG
 TAAAGCTTCTTCTGGACGAGTTGCACCCATCATGGTCCGCCAAGTTTCGATTACTTTGGG
 ACCAGAAATGACACCCACAAGAACTGGACCTGAAGTCATGAATTCACGAATCGGTGGGTA
 AAAACTCTGACCAACCAAGTCTGATAGTGCTGGTCAATCAACTCTTCTGAAAACCTGTG
 AACGAACTCCAATTTTTCGATTGTAAATCCACGTTGTTTCGATGCGCTTTAACACTTCAC
 CCACTAGCCCTCTTTTACACCATCTGGTTTGATGATAAAGAATGTTTGTTCCATACCCG
 TCTCCTTTGTCAGCTTCTTTCTTTTATTTTACCACATCTCGTGGAAAAATGGAGAAAGTT
 TTCAGAAGAGAGAATGAGAGAACCCTCGGGTTCTCTCATCTCTCTTATTCTACTGTTTC
 TTCCACAGTGTCAACGGCAGTATCCACAACACTACTTCTGTTGTTTCTTCATTTCCCTTCTTC
 CTCTACTGGAGGATTAAGGTATTCTTCTTCGTTGACAGCATGTGGTTCAAGGTTACGGTA
 ACGGGCCATACCAGTACCAGCTGGGATGATCTTACCGATGAATAACATTTTCTTTAAAT
 TCCAAGG

ORF Predictions:

ORF #	Start	End	Direction	Length
1	73	453	R	127 aa

>[SEQ ID NO:111] 3859124-1 ORF translation from 73-453, direction R
 VDLQSKNWSFVHRFSEELIDQHYQDLVGQSFYPPIREFMTSGPVLVGVISGPKVIETWRT
 MMGATRPEEALPGTIRGDFAKAAGENEIIQNNVHGS DSEKSQLSREIAPLVL RVDWLNQL
 VKSSF*

Description:

NUCLEOSIDE DIPHOSPHATE KINASE (EC 2.7.4.6) (NDK) (NDP KINASE)
 (ABNORMAL WING DI SCS PROTEIN) (KILLER-OF-PRUNE PROTEIN). -
 DROSOPHILA MELANOGASTER (FRUIT FLY) .

Assembly ID: 3859244

Assembly Length: 578bp

>[SEQ ID NO:25] 3859244 Strep Assembly -- Assembly id#3859244
 ACAACCTAACTACCGNCTAATTCAGCGCGAACTTCTGCAGTAGCTGCTTCAACAACCTTCA
 CGACGTGAAAGGATGAAGCGGTTTTCTTTAGCGTTAACTTCTTTGATTTTAGTATCAAAT
 TCTTGACCTACAAAACGCTCAGCGTTACGTACGAAACGAGTATCCAACATTGAAGCTGGG
 ATAAATCCACGAACACCTTCAAATTCTACTGAAAGTCCACCTTTAACGGCACGCGTTCCCT
 TTAACAGTAACAACCTTCTTCTTCGCGACCAACAAGTTTGTCCCATGCTTTGCGAGCTTCA
 AGGCGTTTTTTTAGATGACAAGGTATGTAAGTGTATCAGTATCTTTACCAACTACTTGACG
 AAGTACAAGAACATCCAATACTTCTCCTACTTTAACAAAGTCATTGATATCTGCATCACG
 ATCGTTTGTCAATTCGCGAAGAGTCAAGACACCCTTCAACACCAGTTCCAGAAGAATGC
 AACGTTAGCTTGAGTCGCATCAACTGTCAATACTTACGACTAACACATCACCAGTCTCA

ACTTGACTNACGCTATTGAGCANATCTTCAAATTCGAT

ORF Predictions:

ORF #	Start	End	Direction	Length
1	310	462	R	51 aa

>[SEQ ID NO:112] 3859244-2 ORF translation from 310-462, direction R
VLKGVLTLLRELTDNRDADINDFVKVGEVLDVLRQVVGKDTDTVTYLVI*

Description:

unknown

Assembly ID: 3859250

Assembly Length: 888bp

>[SEQ ID NO:26] 3859250 Strep Assembly -- Assembly id#3859250
GTAGTTATAGTAGGGGTCGGATTGAAATGCCACNGCGCTTCTTGGAGTTTCTGATACCGT
TTAAATAGCGTTGGGCATTCTGGTTGGGAGTCAGAGCCTTATCAAGCGCAATCATGATA
GGTTGGTTGGTATAGTAGTTGTCTAGGATAACCTGGTTCTTGGTCGTTAGGCACCTGGTG
GAGGAAGGTTGTCAGCAATTCTCCTTTTTGACGAAATTCTTCAGCGTTGTCTGTCGCCAG
TAACTATTTTTCTCTGTTTTTTGAGTTTGTGTCTGGTTTTCTGAAGTTCATTTTCAACACG
ACGAATCAGTTCAGTGGCCTGCTGTTTTGACGCGGTCGCGCTCAGCCTTATCCTTATAGTA
GGTGTCCAACAAATCAGAAAGATTTGCAAAGGCTCTCCACCTGATTTGCAAAGGAAC
TGGACTGAAGGAAGTCTCAGTCAAGCATGGCTTGGTTTCCTGATTGAAAAAATTCGGAA
AGCGGAAAGTTTTTCACTAACCAGTATCCTTTCCAATTCATTTGCCGTATCGCGTCCAG
ACCTTGAAAGAGGCTTTGAAGATTTTTTGCTGTTAGTTCTTGGGTTTGCAGGATTTCAA
GAGCTTTTTCATCCTTGATAGTAAAAGGATTGAGAGATTCTGTACTTGGCGGAGCGATATA
GGTCGATCCTGGAAGTAAGGTGCGGTAGCTATTTTGTGAAAAGCCGACGTGTTTGATAAC
TTCGAGGATTTTATGACTGCTTTTATCCGACCAGTTAGAATATTACTGTGTTTCCCCATA
ATTTGATAATCAAGGTAGCCTGGATATGGTCTCCAATCTCGTTTTTATTGGAAACTGTA
ATTTCCACAATACGGTCATTTTCCACTTGCTCAATCGACTCAATCAGG

ORF Predictions:

ORF #	Start	End	Direction	Length
1	244	402	R	53 aa

>[SEQ ID NO:113] 3859250-1 ORF translation from 244-402, direction R
VGEPFANLSDLLDYYKDKAERDRVKQQASELIRRVENELQKNRHKLKKQEK*

Description:

STRFBP5A NCBI gi: 496253 - Streptococcus pyogenes.
Fibrinogen/Fibronectin binding protein

Assembly ID: 3859588
Assembly Length: 513bp

>[SEQ ID NO:27] 3859588 Strep Assembly -- Assembly id#3859588
ATCGAATTTTGTTCCTTCATAGAGAGCTACCTGAGTTCTATTCAAGCTCAGGTAGTACTT
TCTTATAAACTAGACAACTAACTGTCATTCTACCATCAGATTACAAGACATCATCGTCA
CTCACCTTGGAATTCAATGTCGTACCCCAATGGGTAATTTTACGGTGGGGTTGAGCTAAA
ATTGGTCTGTTTTCATAGATTGTTTGCCATCTATTCCATAGTAGGCCCGTCTTTTCTCA
ATCTTAACTCGCAGATTTCTCATATTTTCTTTGATTGGGAGGTTGAGGACAAAACCTGCA
GTCTGGTTGCGACCGTTTCCTTCCCAAGAATGACTACGAACAACCTTGGTTTCCATCTTTA
TCTACTGGAACCTCTTCCCAAGTTATGGAGTAGCGGGCAATGTAAGCTCCACTGTGTTGA
ATTATCAATGTTTTATCTTTCACAGGGAGTCTGACTGATTGGTTGAACTGGCTTAGAAAC
TTGTGTCGCCGTTTCAGCATTCGTAGCTATAAA

ORF Predictions:

ORF #	Start	End	Direction	Length
1	102	443	R	114 aa

>[SEQ ID NO:114] 3859588-1 ORF translation from 102-443, direction R
VKDKTLIIQHSGAYIARYSITWEEVPVDKDGNOVVRSHSWEGNGRNTAGFVLNLPIKEN
MRNLRVKIEKKTGLLWNRWQTIYENRPILAQPHRKITHWGTTLNSKVSDDDL*

Description:

PNEUMOLYSIN (THIOL-ACTIVATED CYTOLYSIN). - STREPTOCOCCUS PNEUMONIAE.

Assembly ID: 3859774
Assembly Length: 214bp

>[SEQ ID NO:28] 3859774 Strep Assembly -- Assembly id#3859774
ATCGAATTCTAACATGTGCTTCTCCTTCTATTGTTCCCTATCTTTAAAATCTACTCCTTCA
TGCTCCAAGAGCCAAGCTTTCTTTTCCACTCCTGCAGCATAACCTGTCAGACGCTTGCCT
GCTCCCAACACACGATGACAAGGTACTAGGATAGACCAAGGATTGCGTCCCCTGCTCCA
CCAATTGCTTGAGCAGAAGCCACTTGCAGGTCTT

ORF Predictions:

ORF #	Start	End	Direction	Length
1	9	131	R	41 aa

>[SEQ ID NO:115] 3859774-1 ORF translation from 9-131, direction R
VLGAGKRLTGYAAGVEKKAWLLEHEGVDFKDRNNRRRSTC*

Description:

GLUTAMATE RACEMASE (EC 5.1.1.3). - ESCHERICHIA COLI.

Assembly ID: 3860140

Assembly Length: 1084bp

>[SEQ ID NO:29] 3860140 Strep Assembly -- Assembly id#3860140
CTCCAGCAATGGATCCAAGTATGATGGGCGGGATGATGTAAGCTTTCTATAGAAAACACC
TTATAAAAAACACGAAAGGAGGGAATGACTAACCCTTCTTTTTTATAATATTTACTTCTAA
GATTGATGGTGAGCTCTCCTAAGTATATGATAAAATAAGACTAGAGGAAAGGAGAAGAA
CATGATCGATGTACAAGAAATTCTGTGCAAGATGACCCCAATCAGAAGATTAATTATGA
CCGTGTCATGCAGAAAATGGTACAAGCATGGGAAAAAAATGAGTAGCGGCCAACCATTCT
CGTGCATGTTTGGCTGTGCCCCCTGTAGTACCTATACACTAGAATATTTGACCAAGTATGC
AGATGTGACCATCTATTTTGCCAATTCTAATATCCATCCCAAGGCAGAATACCATAAGCG
GGTCTATGTACCAAGAAATTTGTTAGTGATTTTAATGAGCAGACAGGAAATACGGTTCA
GTACCTAGAAGCTCCCTACGAACCCAATTAATACCGAAAACCTAGTTAGGGGGCTAGAGGA
GGAGCCCGAAGGTGGCGACCGTTGCAAGGTTTGTGTTTACTACCGACTGGATAAAACAGC
GCAAGTGGCTATGGACTTGGGCTTTGACTACTTTGGTTTCAGCCTTGACCATCAGTCCTCA
TAAGAATTCTCAAACCTATCAATAGCATCGGAATCGATGTGCAAAAAATTTACACGCCCCA
CTATCTTCCCAACGATTTCAAGAAAAATCAAGGCTACAAACGTTTCAGTAGAGATGCGTGA
GGAGTATGATATCTATCGTCAATGTTATTGTGGCTGCGTCTATGCAGCCCAAGCCCAGAA
TATTGACCTGGTTTAAGTTGAGTAGGACGCCACAGCATGCTTGCTGGATAAGGATGTTGA
GAAAGACTATTCTCATATCACATTTATAGTAGATTGAACTAGAATAGTACACCTTTACT
TCTCAAACATTGTTAGAAATCGATTTCGGCTGTCCTTATTTTCATTTTAATATACTGGTACG
AAATTAGATATATCAATGATAACTTGCCTCAAGGTAGGTTTTTTGATAGTAGAAAAGCGA
TAGA

ORF Predictions:

ORF #	Start	End	Direction	Length
1	302	511	F	70 aa
2	605	856	F	84 aa

>[SEQ ID NO:116] 3860140-1 ORF translation from 302-511, direction F
VHVCCAPCSTYLTLEYLTKYADVITYFANSNIHPKAEYHKRVYVTKKFVSDFNEQTGNTVQ
YLEAPYEPN*

Description:

unknown

>[SEQ ID NO:117] 3860140-2 ORF translation from 605-856, direction F
 VAMD LGFDYFGSALTISPHKNSQTINSIGIDVQKIYTPHYLPNDFKKNQGYKRSVEMREE
 YDIYRQCYCGCVYAAQAQNIDL*

Description:

unknown

Assembly ID: 3860206

Assembly Length: 1124bp

>[SEQ ID NO:30] 3860206 Strep Assembly -- Assembly id#3860206
 ATCGAATTCATTGACTGCCTGAAAAGACTTCAACTCGTCTGCCTGATAACCGAAAGACTT
 GGTTACTTTTGATACCTGATACGGACTCCTGTACCTTGTTATTGAGTTCAGAAAAAGCAGC
 TTGGGATTCGCCAAAGGCCTTATGAGTCTTTCTCCCTAGGCGACTAGTCGTATAGGCCAT
 GAAAGGTAGGGGGAGAATGGCAACAAGAGTCATCTGCCATGAGATGCTAAAGAGCATGGT
 CAACAAAGTCACCAGAGCCGTGATAGAGGCATCCACCGCAGACATGACACCGCCACCTGC
 TAAACGAGTCAAGGAATTGATATCATTGGTTGCGTGTGCCATCAGATCACCCGTCCGATA
 GGTTTGATAAAAGGCTGACGACATTTTTGTGAAATGCTTAAACAAGCGAGACCGCATGAT
 CTGTCCCAAGCAATAAGAGGTCCCAAGGATATACATACGCCACACATAGCGCAAATAGTA
 CATACCAAAGGCTGCAAGTAGCAAGTAAAATAGGCTAAGAAGGAGGTCCTGCTGGGTAA
 TTGCCCCGATGTGATGGCATCAATAACCCGCCCATTAACCATAGGAGGAATGAGATTGAG
 GACGGAAACCAAGACCAGGGCCACAATCCCGACTAGATAACGGCGTTTTTCTAAGTTGAA
 AAACCACCAAAATTTTTGAATAATGGACATAAAATCCCTTTCTGGATTGCAAATAGAAAC
 CTGAGGCCAATACTCAATGGAAAATCAAAGAGCAAAC TAGGAACTAGCCGCAGGCTGCT
 CAAAGCACTGCTTTGAGGTTGTAGATAGAACTGACGAAGTCAGTAACCTACATACGGCAA
 GGCGACGTTGACGCCGTTTGAAGAAATTTCCGAAGAATACAAGACCCAGGTTTTTCTTA
 TTTATAAGTTACCACTGTAAACAGCACCCCTTGTCATATTCAGCAATAAAGATATTGGCTAC
 ATTGTCATGCCCTTGTTTACTGAGGTTATCAAGCAACCACTCCTCGCTACGAACAATCGA
 TCCCAAGACATCTACTTGAATCACACCGTCAGTCACAACTGGATACTTAGGATTTTCATC
 TCCCATTTGCACAACGATGAGTTGCCCATTTTGCTCTTGACACAG

ORF Predictions:

ORF #	Start	End	Direction	Length
1	898	1056	R	53 aa

>[SEQ ID NO:118] 3860206-2 ORF translation from 898-1056, direction R
 VTDGVIQVDVLGSIVRSEEWLLDNLSKQGHNDVANIFIAEYDKGAVTVVITYK*

Description:

unknown

Assembly ID: 3860270

Assembly Length: 1242bp

```
>[SEQ ID NO:31] 3860270 Strep Assembly -- Assembly id#3860270
TTACCTTCATTGCAGCCATTATTGGTTCTTGTGTCAGCCAGATTTTAAGTATTCTTTATA
AGACACCTGCTGTGGTCTTTATCTTGGCCATTTTGGCACCGCTGGTTCCAGGTTATCTCT
CCTACCGAACAACTGCCTTTTTTTGTGACAGGGGACTATAATAAAGCACTGGCAAGTGCGA
CCTTGGTTGTCATGTTGGCTTTGGTAATCTCTATTGGAATGGCTAGCGGAACAGTGATTC
TCAGACTGTATCATTATATAAAAACACATCGAGTATCGTAGACTTTACAGAAATAAAAGA
ATTTTCTGAAAAATGAGATAAAATAAATAACAACGCTTTCTATATGTGCGAGAATACCGC
ACTTATGAAGAAATTGCGGCTGATTTTGGTATCCACGAAAGCAACTTAATCCGTCGGAGC
CAATGGGTTGAAGTAACTCTTGTTCAAAGTGGTGTACGATTTCAAAACTCATCTTAGT
GCTGAGAATACGGTGATTGTGGATGCAACAGAGGTAAAAATCAATCGCCCTAAAAAACAA
TTAGCGAATGATTCTGGTAAAAAGAAATTTACGCTATGAAGGCTCAGGCGATTGTCACA
AGTCAAGGGAGAATTGTTTCTTTGGATATCGCTGTGAACTATTGTCATGATATGAAGTTG
TTCAAAATGAGTCGCAGAAATATCGGACAAGCTGGAAAAATCTTGGCTGATAGTGGTTAT
CAAGGGCCCCATGAAGATATATCCTCAAGCACAACTCCACGTAAATCCAGCAAACCTCAAG
CCGCTAATAGCTGAAGATAAAGCTTATAACCATGCGCTATCCAAGGAGAGAAGCAAGGTT
GAGAACATCTTTGCCAAAGTAAAAACGTTTAAATGTTTTCAACAACCTATCGAAATCAT
CGTAAACGCTTCGGATTACGAATGAATTTGATTGCTGGCATTATCAATTATGAACTAGGA
TTCTAGTTTTTGCAGGAAGTCTATTATTTTCTTATTGTCTGTAAGTCTACTGACCTTGT
GTTTATCCCAGTCATGGTTTCTAGTTCGGGCTCAGAGTTTCAAAGTGGATGGCAAGAGCA
TCAATTGATTGCTGAGAAGGTTAGTAAAACACTTGACAAGACATTTGATAAGGATGTCAG
AAAAATTCCGACCAGTCAGTTTTATCAAAAATTTGTAGATGAGATGGGAAGGATTTACTC
AGGAAATTTGATCCTCCCAGGAGCTGATAACTGTGAATGGAG
```

ORF Predictions:

ORF #	Start	End	Direction	Length
1	346	966	F	207 aa

```
>[SEQ ID NO:119] 3860270-1 ORF translation from 346-966, direction F
VREYRTYEEIAADFGIHESNLIRRSQWVEVTLVQSGVTISKTHLSAENTVIVDATEVKIN
RPPKQLANDSGKKKFHAMKAQAIVTSQGRIVSLDIAVNYCHDMKLFKMSRRNIGQAGKIL
ADSGYQGPMKIYPQAQTPRKSSKLPLIAEDKAYNHLSKERSKVENIFAKVKTFKMFST
TYRNHRKRFGRLRMNLIAGIINYELGF*
```

Description:

ISL2 protein - *Lactobacillus helveticus* (Probable transposase)

Assembly ID: 3860438

Assembly Length: 1575bp

```
>[SEQ ID NO:32] 3860438 Strep Assembly -- Assembly id#3860438
```

GTGATGGGGCCTCAGGGAAATGGTTTTGACTTGTCTGACCTTGATGAGCAGAATCAGGTT
 CTCCTTGTGGTGGTGGGATTGGTGTTCACCCTTGCTTGAGGTGGCCAAGGAATTGCAT
 GAACGTGGAGTGAAAGTAGTGACAGTCCTCGGTTTTGCTAATAAGGATGCTGTTATTTTG
 AAAACGGAATTGGCTCAGTATGGTCAGGTCCTTTGTAACGACAGATGATGGTTCCTTATGGC
 ATCAAGGGAAATGTTCCGTTGTTATCAATGATTTAGATAGTCAGTTTGATGCTGTTTACT
 CGTGTGGGGCTCCAGGAATGATGAAGTATATCAATCAAACCTTTGATGATCACCCAAGAG
 CCTATTTATCTCTGGAATCTCGTATGGCTTGTGGGATGGGAGCTTGCTATGCCTGTGTTC
 TAAAGTACCAGAAAGCGAGACGGTCAGCCAACGCGTCTGTGAAGATGGTCCTGTTTTCC
 GCACAGGAACAGTTGTATTATAAGGAGAAAATTATGACTACAAATCGATTACAAGTGTCT
 CTACCTGGTTTTGGATTGAAAAATCCGATTATTCCAGCATCAGGCTGTTTTGGCTTTGGA
 CAAGAGTATGCCAAGTACTATGATTTAGACCTTTTAGGTTCTATTATGATCAAGGCGACA
 ACCCTTGAACCACGTTTTTGGGAATCCAACCTCCAAGAGTGGCAGAGACGCCTGCTGGTATG
 CTCAATGCAATTGGCTTGCAAAATCCTGGTTTTAGAGGTTGTTTTGGCTGAAAAGCTACCT
 TGGCTGGAAAGAGAATATCCAAATCTTCCTATTATTGCCAATGTAGCTGGTTTTTCAAAA
 CAAGAGTATGCAGCTGTTTTCTCATGGGATTTCCAAGGCAACTAATATAAAAGCTATCGAG
 CTCAATATTTCTTGTCCCAATGTTGACCACTGTAATCATGGACTTTTGATTGGTCAAGAT
 CCAGATTTGGCTTATGATGTGGTGAAAGCAGCTGTGGAAGCCTCAGAAGTGCCAGTTTAT
 GTCAAATTAACCCCGAGTGTGACCGATATCGTTACTGTCGCAAAGCTGCAGAAGATGCG
 GGAGCAAGTGGCTTGACTATGATCATACTCTGGTGGGATGCGCTTTGACCTCAAACCAG
 AAAACCAATCTTGGCCAATGGAACAGGTGGAATGTCAGGTCCAGCAGTTTTCCAGTAGCC
 CTCAAATCATCCGCCAAGTAGCCCAAACAACAGACCTGCCTATCATTGGAATGGGGGGA
 GTGGATTCGGCTGAAGCTGCCCTAGAAATGTATCTGGCTGGGGCATCTGCTATCGGAGTT
 GGAACAGCTAACTTTACCAATCCTTATGCCTGCCCTGACATCATCGAAAATTTACCAAAA
 GTCATGGATAAATACGGTATTAGCAGTCTGGAAGAACTCCGTCAGGAAGTAAAGAGTCT
 CTGAGGTAACTGCAATCAATCTGTTCTTGATTTTTTATTAGTTTGTAATATGAATTTAG
 GAGAATTTTGGTACAATAAAATAAATAAGAACAGAGGAAGAAGGTTAATGAAGAAAGTAA
 GATTTATTTTTTTAG

ORF Predictions:

ORF #	Start	End	Direction	Length
1	1	276	F	92 aa
2	460	1128	F	223 aa

>[SEQ ID NO:120] 3860438-1 ORF translation from 1-276, direction F
 VMGPQGNQFDLSDLDEQNQVLLVGGGIGVPPLEVAKELHERGVKVVTVLGFANKDAVIL
 KTELAQYQGQVFTTDDGSYGIKGNVPLLSMI*

Description:

unknown

>[SEQ ID NO:121] 3860438-3 ORF translation from 460-1128, direction F
 VKMVLFSAQEQLYYKEKIMTTNRLQVSLPGLDLKNPIIPASGCFGFGQYAKYYDLDLLG

SIMIKATTLEPRFGNPTPRVAETPAGMLNAIGLQNPGLLEVLAEKLPWLEREYPNLPIIA
 NVAGFSKQEYAAVSHGISKATNIKAIELNISCPNVDHCNHGLLIGQDPDLAYDVVKAAVE
 ASEVPVYVKLTPSVTDIVTVAKAAEDAGASGLTMIILWWDAL*

Description:

DIHYDROOROTATE DEHYDROGENASE (EC 1.3.3.1) (DIHYDROOROTATE OXIDASE)
 (DHODEHASE). - BACILLUS SUBTILIS.

Assembly ID: 3860544

Assembly Length: 776bp

>[SEQ ID NO:33] 3860544 Strep Assembly -- Assembly id#3860544
 CTAAGATATCAGAATAACAACGAAATCGAAGCATTAACAAATATTACTTCTAAGAAT
 AGCGAGATTGATAGTCAACAAAGCAATATTAAGGATATGACCGTACCTATAATGATCCAA
 CTTCTCAGGCTTATAATATTTATGCTCAATTAATTAGTGAGTTAGGTACTGCTCGTTCAA
 ACAACAATAAAAGTATTACAGAGCTTGAGGCTAATCTTGGAGTGGCAACAGGTCAAGATA
 AAGCTCATAGTATATTAGCGTCAAATGAAGGTACTCTGCATTATCTGGTACCTTTGAAAC
 AAGGAATGTCTATTTCAGCAGGGGCAAACGATAGCAGAAGTTTCAGGGAAAGAAAAGGTT
 ACTATGTAGAGGCTTTTGTACTTGCGAGTGATATTTCTCGTGTTTCAAAGGAGCAAAAG
 TTGATGTTGCTATTACTGGTGTGAATAGTCAAAAATATGGAACACTAAAGGGACAAGTCA
 GACAGATTGATTTCAGGAACAATTTCCCAAGAAACGAAAGAGGGGAATATTAGCCTCTATA
 AAGTCATGATAGAATTAGAAACCTTAACCTCTAAAACATGGAAGCGAGACGGTCATACTCC
 AAAAGGATATGCCAGTTGAAGTGCGGATTGTCTATGATAAAGAAACCTATCTTGATTGGA
 TTTTAGAAATGTTAAGTTTCAAGCAATAATTGGTTTTAAACCTTAGGTAACCTATAAAAA
 CAAATAAGGTAGAGAAAGGATATTTTATCTAAGTTAGCTCACATTACTGCCATTCC

ORF Predictions:

ORF #	Start	End	Direction	Length
1	222	689	F	156 aa

>[SEQ ID NO:122] 3860544-1 ORF translation from 222-689, direction F
 VATGQDKAHSILASNEGTLHYLVPLKQGMISIQGQTIAEVSGKEKGYEAFVLASDISR
 VSKGAKVDVAITGVNSQKYGTGKQVRQIDSGTISQETKEGNISLYKVMIELETTLTKHG
 SETVILQKDMPVEVRIVYDKETYLDWILEMLSFKQ*

Description:

unknown

Assembly ID: 3860558

Assembly Length: 1487bp

>[SEQ ID NO:34] 3860558 Strep Assembly -- Assembly id#3860558

60

CTGGCCTTTCTCCACCAAAATTGTTTCCTTGAGGGAAGGAAGTCAGAACACTAGCCGTTGC
 ATCTTCCTTTTGGCTTTTCAATCGTAATTCCAGATAATTTTTCCCATTCCTTTTGGTGACC
 CCGGGAGGCAGGATTGAATGGCTTGAGGGAAATGACAACTTGTCCTAGCAAGAATGGTC
 AAGGCACCTCCGTCTACAATCAAAATCTGATTTGGGCTTAAATTAACAAAGACCTGTTTT
 ACTAGATTTTCTCCAGAAGCATCGTCTCGTAAACCAGGCCCCAGCAAGATAACTTCTGCC
 TTCTCCAATTGCTCTTTTAACAATTGCTGGTCTTGAAGAGAAAAGGCCATAGGCTCAGGT
 AAATGGCTGTGCAGAGCCGGGATATTTTCCCTGTCCGTTCCAACGGTCACCAATCCTGCA
 CCGCTTTTTTACAGCTGCTAAAGCAGCCATGATGATGGCACCTCCATAAGGATAAGTACCA
 CCAAGCAGCAGCAGACGACCATAATCTCCTTTATGACTTGAACGAGAACGTTCAATAATA
 ACTTTTTTCTAGTAAGGTTTGATTAATCACTTTCATCCTTTTTTCCCTCTCACTTTTTATTAT
 ACAACAAAAAGGAGACGCAGACCTCCTTTTGTAATCTTATATCTAAAATTTAATATTCAT
 TTCTGCCATTTTAGATATAGCTATAGAAAAATACACTCTATTAATCGAATGTTTCTCTTAT
 TTTCTATCCAATGTCCGAAGTGCTGCTTGATAAGTTTGCTCCATCAGCATGGTAATGGTC
 ATAGGACCGACACCTCCAGGGACTGGCGTGATATGGCTAGCAAGTGGTGCAACTGCCTCA
 TAATCAACATCTCCACAGAGCTTCCCATTTTCATCTCGGTTTCATCCCAACGTCAATGACA
 ACCGCACCTGGTTTGGACAAAGTCAGCAGTCACAACTTGGCGCGGCCGATTGCGACTACA
 AGAATATCTGCTTTAGCAGCCACCTTGGCAAGATTATGAGTTCGTGAGTGGGCCAAGGTT
 ACTGTGCGCATTTTTAGCCAAAAGAAGCTGAGCCATAGGTTTTTCCAACGATATTTGAACGA
 CCGATTACGACCGCATTTTTACCTTCCAAGTCAATCCCATATTCATGAAACATTTCCATA
 ATTCCTGCAGGTGTGCGAGGGAATCATGACTGGATGTCCAGACCAAAGACGTCCCATGTTT
 AGGGGATGGAAACCATCCACATCCTTTTTCTGGGTCAATGGCTAATAAAACCGCCTCTTCA
 TCGATATGTTTTGGTAATGGCAACTGGACCAAATCCCATGCCAAGCTGGATCCTGATTA
 TATTTAGCAATCAGGTCTAACAATTCCTCTTGAGTAATGGTCTCTGGAACCTCGCACTACT
 TCGGTACGGGAACCAGCCGCAAGAGCTGACCTCTCCTTGTTGCGAACGTTAACTTGGCT
 GGCTGGATTATCCCCAACCAAAATCACTACCAAAACCAGGCACTAGAG

ORF Predictions:

ORF #	Start	End	Direction	Length
1	717	1376	R	220 aa

>[SEQ ID NO:123] 3860558-2 ORF translation from 717-1376, direction R
 VRVPETITQEELLDLIAKYNQDPAWHGILVQLPLPKHIDEEAVLLAIDPEKDVDGFHPLN
 MGRLWSGHPVMIPSTPAGIMEMFHEYGIDLEGKNAVVIGRSNIVGKPMQQLLAKNATVT
 LAHSRTHNLAKVAAKADILVVAIGRAKFVTADFVKPGAVVIDVGMNRDENGKLCGDVDYE
 AVAPLASHITPVPGGVGPMTITMLMEQTYQAALRTLDRK*

Description:

5,10-methylene-tetrahydrofolate dehydrogenase (fold) homolog -
 Haemophilus infl uenzae (strain Rd KW20)

Assembly ID: 3860568

Assembly Length: 1634bp

>[SEQ ID NO:35] 3860568 Strep Assembly -- Assembly id#3860568
 CGTGCCTTGGCCAATGATCCAAAAATCTTGATTTTCAGACGAGTCGCTTCAAATTTTCGGCC
 CCTGGACCCCTTAAGACCAACCCAAGCAGATTTTGGCCCTTGGTTGCAAGATTTGAACCAA
 AAATTAGGCTTGACTGTTGTCTCTGATTACGCATGAAATGCAGATTGTCAAAGACATTGCC
 AACCGTGTGTCAGTTATGCAGGATGGGCATTTGATTGAAGAGAGTAGTGTGCTTGAAATC
 TTCTCAGACCCTAAACAACCTTTGACTCAAGACTTTATCTCAACAGCTACAGGTATTGAC
 GAAGCCATGGTCAAAATCGAGAAGCAAGAAATCGTGGAACACTTGTCTGAAAACAGTCTC
 TTGGTGCAACTCAAGTACGCTGGATCTTCAACAGACGAGCCACTTTTGAATGAATTGTAC
 AAGCATTATCAAGTAATGGCTAATATTCTCTATGGGAATATCGAAATCCTCGATGGTACT
 CCTGTTGGAGAATTGGTGGTGGTCTTGTCTAGGTGAAAAAGCAGCGCTGGCAGGTGCTCAA
 GAAGCCATTCGTCAAGCAGGCGTACAGTTAAAAGTATTGAAGGGAGGACAGTAAGATGGA
 ATCATTGATTCAAACCTATTTACCAAATGTCTATAAGATGGGTTGGTCTGGTCAGGCAGG
 CTGGGGAACAGCTATCTACCTAACCTCTATATGACAGTTCTTTCTTCATTATCGGAGG
 CTTCTTGGGGCTAGTGGCAGGTCTCTTTCTCGTCTTGACAGCGCCAGGTGGTGTCTTGGA
 GAATAAAGTCGTATTCTGGATTTTAGACAAAATTACCTCAATTTTTCGTGCGGTTCCCTT
 TATCATCCTCTTGGCAATCTTGTCAACCACTTTCTCACTTGATTGAAAAACAAGTATCGG
 GCCAAATGCAAGCCCTTGTCCCACTTTCTTTTGCAGTCTTTGCCTTCTTTGCCCGTCAGG
 TGCAGGTGTCTTGGCTGAAATGGATGGCGGTGTCTTATGAGGCGGGCTCAAAGCGAGCGG
 AGCGACTTTCTGGGACATCGTGGGTGTTTACCTATCAGAAGGTCTTCCAGATTTGATCCG
 TGTGACGACTGTGACCTTGATTTCCCTTGTGGGGAAACAGCTATGGCCGGTGCGGTGG
 AGCTGGTGGTATCGGTAACGTAGCCATCGCTTATGGATTTAACCGCTACAATCACGATGT
 GACCATCTTGGCAACCATCGTTATCATTTTGATTATCTTTGCAATCCAATTCTTAGGAGA
 TTTCTTGACTAAGAAATTGAGCCATAAATAAAAAAGAGCCGTGTGGCTCTTTTAACTGA
 TCAGATTTTCTGGGCAAATTTTTTACTCAAGGCTTGTCCAATCAAGGCACCCACTAGGGC
 TCCGATGACAATACTTGGGATAAATAGAAGGACAGTTCCAGGGTTTGGAGCGACCATGAT
 GCGGTGATATATTCTTGGGATTTTCTCTTGGCAGAAGAGTAGCCATATAGGCTTTGGG
 CGCAATCCACATAAGCAAGATTGGTCTGTGTACTAAAGGCGAAAATAATGAAAGAAAG
 GAAGTTCTTTGTTTTGTCCTTGTATTTTCTTAAATGAGCTACTCCATCTGCTAGGAGGCC
 ACAGATAATTCGAT

ORF Predictions:

ORF #	Start	End	Direction	Length
1	1040	1291	F	84 aa

>[SEQ ID NO:124] 3860568-3 ORF translation from 1040-1291, direction F
 VGVYLSEGLPDLIRVTTVTLLISLVGETAMAGAVGAGGIGNVAIAYGFNRYNHDVTILATI
 VIILIIFAIQFLGDFLTKKLSHK*

Description:

unknown

Assembly ID: 3860582
 Assembly Length: 1087bp

```
>[SEQ ID NO:36] 3860582 Strep Assembly -- Assembly id#3860582
GGAATCATGATGATGTCACTGCTAAATGGTTTCTTAGAAAAAATATTTCCCTGAGCGCTTA
CAGATTAGTTTGGGCTTGCTGATTTTATCATTGAGCGGTACAGCTCCCTTCTGGTACCAA
GCCTATCCCTTTGTCTTTGGAACACGGCTTCTCTTTGGTTTGGGTCTTGGGATGATCAAT
GCCAAGGCCATTTCTATTATCAGTGAACGCTACCAAGGAAAAAGGCGAATTCAGATGTTA
GGGCTACGCGCTTCTGCAGAGGTCGTTGGAGCTTCTCTCATTACCTTGGCCGTCGGTCAA
GTTGTTGGCCTTTGGTTGGACAGCTATCTTTCTAGCCTATAGTGCTGGATTTTTTGGTGCT
GCCCCCTTATCTGCTCTTTGTCCCTTATGGAAAATCAAAGAAAAGAAGTCAAGAAAAGAGC
GAAGGAAGCAAGTCGTTTAACTCGAGAAATGAAAGGCTTGATTTTTTACCTTAGCTATCGA
AGCGGCAGTTGTAGTTTGTACCAATACAGCTATTACCATCCGTATTCCAAGTTTGATGGT
GGAAAGAGGATTGGGGGATGCCAGTTATCTAGTTTGTCTTAGTATCATGCAGTTGAT
CGGGATTGTGGCTGGGGTGAGTTTTTCTTCTTGATTTCTATCTTTAAAGAGAAACTGCT
CCTCTGGTCTGGTATTACCTTTGGCTTGGGGCAAATCGTGATTGCCTTGTCTTCATCCTT
GTGGGTGGTAGTAGCAGGAAGTGTTCTGGCTGGATTTGCCTATAGTGAGTCTTGACGAC
GGTCTTTCAACTTGTCTCTGAACGAATTCAGCTAAACTCCTCAATCAAGCAACTTCATT
TGCTGTATTAGGCTGTAGTTTCGGAGCCTTTACGACCCCATTCGTTCTAGGTGCAATTGG
CTTACTAACTCACAATGGGATGTTGGTCTTTAGTATCTTAGGAGGTTGGTTGATTGTAAT
CTCTATCTTTGTCATGTACCTACTTCAGAAGAGAGCTCTAGGATTGATTCCTAAGTTTTT
CTTTTGATACTCAATGAAAATCAAAGAGCAAACCTATAGTTGATTGAGTTTGAATAGTAT
GCTGTAG
```

ORF Predictions:

ORF #	Start	End	Direction	Length
1	356	1027	F	224 aa

```
>[SEQ ID NO:125] 3860582-1 ORF translation from 356-1027, direction F
VLPLYLLFVPYGKSKKEVKKRAKEASRLTREMKGLIFTLAIEAAVVVCTNTAITIRIPSL
MVERGLGDAQLSSFVLSIMQLIGIVAGVSFSFLISIFKEKLLLWSGITFGLGQIVIALSS
SLWVVVAGSVLAGFAYSVVLTTVFQLVSEIPAKLLNQATSFVAVLGCSFGAFTTPFVLGA
IGLLTHNGMLVFSILGGWLIVISIFVMYLLQKRALGLIPKFFF*
```

Description:
 unknown

Assembly ID: 3860724
 Assembly Length: 1191bp

```
>[SEQ ID NO:37] 3860724 Strep Assembly -- Assembly id#3860724
GGATTCCAACGATTATGAACTTGACTGGTCCACTGATTCATCCAATGGCTTTAGAAACAC
```

AGCTTTCTTGGAATTAGTCGTCCAGACTCCTAGAAAGTACAGCTCAGGTTTTGAAAATAT
 GGTCCGAAACGTGCCATCGTGGTTGCTGGACCAGAAGGGTTGGATGAAGCTGGCTTGAAC
 GGAACAACCNAGATTGCACTTNTTGA AAAATGGCGAAATCAGCTTGTCAAGCTTTACTCCA
 GAGGATTTGGGAATGGAAGGCTATGCTATGGAAGATATTCGTGGTGGGAATGCTCAGGAA
 AATGCAGAAATTTTGCTTAGCGTTCTGAAAAACGAAGCAAGTCCATTCTTGAAACGACA
 GTCTTGAATGCTGGTCTTGGTTTCTATGCTAATGGTAAGATTGATAGCATCAAGGAAGGA
 GTTGCCCTTGGCCCGTCAAGTGATTGCTAGAGGCAAGGCCCTTGAAAACTCAGACTGTTA
 CAGGAGTACCAAAAATGAGTCAGGAATTTTTAGCACGAATCTTAGAGCAGAAGGCGCGTG
 AGGTGGAGCAGATGAAGCTGGAGCAAATCCAGCCTCTGCGCCAGACCTATCGCTTGGCAG
 AATTTTTGAAGAATCATCAGGACCGCTTGCAGGTAATCGCTGAGTCAAGAAAGCTAGCCC
 TAGTTTGGGAGATATCAATCTCGATGTGGATATTGTGCAACAGGCCCAGACTTATGAAGA
 AAACGGAGCAGTGATGATTTTCGGTGTGACAGATGAGGTTTTCTTTAAAGGGCATTGTGA
 TTATCTACGGGAAATTTCCAGTCAGGTAGAGATTCCGACGCTCAACAAAGACTTTATCAT
 AGATGAAAAGCAAATCATCCGCGCTCGCAATGCAGGTGCGACAGTTATCTTGCTTATTGT
 GGCAGCCTTGTCCGAAGAACGCCTCAAGGAAGTGTATGACTACGCGACAGAGCTTGGTCT
 GGAAGTCTTAGTGAGACTCACAACTAGCTGAACTAGAGGTAGCCACAGACTTGGTGG
 CTGAGATTATCGGGGTCAACAACCGCAACTTGACTACCTTTGAAGTCGACTTGCAGACCA
 GTGTAGATTTAGCCCCTTACTTTGAGGAAGGTCGCTATTACATTTCTGAATCTGCCATTT
 TCACAGGGCAGGATGCGGAACGACTAGCCCCATACTTTAACGGAATTCGAT

ORF Predictions:

ORF #	Start	End	Direction	Length
1	139	498	F	120 aa
2	686	1024	F	113 aa

>[SEQ ID NO:126] 3860724-1 ORF translation from 139-498, direction F
 VVAGPEGLDEAGLNGTTXIALXENGEISLSSFTPEDLGMEGYAMEDIRGGNAQENAEILL
 SVLKNEASPFLETTVLNAGLGFYANGKIDSIKEGVALARQVIARGKALEKLRLQLQEYQK*

Description:

ANTHRANILATE PHOSPHORIBOSYLTRANSFERASE (EC 2.4.2.18). - LACTOCOCCUS
 LACTIS (SUB SP. LACTIS) (STREPTOCOCCUS LACTIS).

>[SEQ ID NO:127] 3860724-2 ORF translation from 686-1024, direction F
 VDIVQQAQTYEENGAVMISVLTDEVFFKGHLDDLREISSQVEIPTLNKDFIIDEKQIIRA
 RNAGATVILLIVAALSEERLKLKYDYATELGLEVLVETHNLAELEVAHRLGG*

Description:

INDOLE-3-GLYCEROL PHOSPHATE SYNTHASE (EC 4.1.1.48) (IGPS). -
 LACTOCOCCUS LACTIS (SUBSP. LACTIS) (STREPTOCOCCUS LACTIS).

Assembly ID: 3860858

Assembly Length: 858bp

```
>[SEQ ID NO:38] 3860858 Strep Assembly -- Assembly id#3860858
ATCGAATTTGCCAACCAAGAAAAATATCCCTTGGATGGTTCTTGGCAATGCAAGCAATAT
CATCGTTCGTGATGGTGGGATTCGTGGATTTGTCATCTTGTGTGACAAGCTCAATAACGT
TTCTGTTGATGGCTATAACCATTTGAAGCAGAAGCTGGGGCTAACTTGATTGAAACAACTCG
CATTGCCCTCCGTCATAGTTTAACTGGCTTTGAGTTTGCTTGTGGTATTCCAGGAAGCGT
TGGCGGTGCTGTCTTTATGAATGCGGGTGCCTATGGTGGCGAGATTGCTCACATCTTGCA
GTCTTGTAAGGTCTTGACCAAGGATGGAGAAATCGAAACCCTGTCTGCTAAAGACTTGGC
TTTTGGTTACCGCCATTCAGCTATTCAGGAGTCTGGTGCAGTTGTCTTGTGAGTTAAATT
TGCCCTAGCTCCAGGAACCCATCAGGTTATCAAGCAGGAAATGGACCGCTTGACGCACCT
ACGTGAACTCAAGCAACCTTTGGAATACCCATCTTGTGGCTCGGTCTTTAAGCGTCCAGT
CGGGCATTTTGCAGGTCAGTTCGAATTTCAGAAGCTGGCTTGAAAGGCTATCGTATCGGT
GGCGTAGAAGTGTGAGAAAAGCATGCAGGATTTATGATCAATGTCGCAGATGGAACGGCC
AAAGACTACGAGGACTTGATCCAATCGGTTATCGAAAAAGTCAAGGAACACTCAGGTATT
ACGCTTGAAAGAGAAGTCCGGATCTTGGGTGAAAGCCTATCGGTAGCGAAGATGTATGCA
GGTGGTTTTACTCCCTGCAAGAGGTAGTGGGGACCTGACAGAGCCCCGATCGGTTAATCT
ATGAAAAAGAAGGAATTT
```

ORF Predictions:

ORF #	Start	End	Direction	Length
1	610	807	F	66 aa

```
>[SEQ ID NO:128] 3860858-1 ORF translation from 610-807, direction F
VSEKHAGFMINVADGTAKDYEDLIQSVIEKVKEHSGITLEREVRLGESLSVAKMYAGGF
TPCKR*
```

Description:

unknown

Assembly ID: 3860890

Assembly Length: 980bp

```
>[SEQ ID NO:39] 3860890 Strep Assembly -- Assembly id#3860890
CTGAAAAAACAGGTTTTGACTATGNAGATTGACAGACGACCGTTTCGGAGGTGCAGATATT
GATGCAGCAGGACCTCCCTTACCTGATGAAACCCTTAAGGCAAGTAGGGAAGCAGATGCT
ATCCTACTAGTAGCTATCGGTAGTCCTCAGTATGATGGAGTAGCGGTTTCGCCCTGAACAA
GGCCTGATGGCTCTCCGTAAGAACTCAATCTTTACGCTAATATTCGTCCTGTAAAAATCT
TTGACAGTCTCAAGTATTTGTCACCACTCAAACCGGAACGAATTTCTGGTGTAGACTTCG
TCGTGGTGCGTGAATTGACTAGGCGAGATTTACTTTGGAGATCATATCCTTGAAGAGCGC
AAAGCGCGTGATATCAACGACTATAGCTATGAGGAAGTGGAGCGGATTATTTCGCAAAGCC
TTTGCCATCGAATTGCAAGAAATCGCAGAAAAATCGTTACTAGTATCGATAAGCAAAATG
```

TTCTAGCGACCTCAAAACTCTGGCGGAAAGTAGCTGAGGAAGTCGCACAGGATTTCTCAG
 ATGTAACCTTGGAACACCAGCTGGTAGACTCAGCTGCTATGCTTATGATTACCAATCCTG
 CTAAGTTTGATGTTATTGTAACGGAGAATCTTTTTGGAGATATTTTATCTGATGAATCAA
 GCGTCTTATCTGGTACACTTGGGGTTATGCCATCAGCCAGTCATTCTGAAAATGGACCAA
 GTCTCTATGAACCTATTACGGTTTCAGCACCTGATATTGCAGGTCAAGGAATTGCCAATC
 CTATTTCCATGATTTTATCAGTTGTCATGATGTTGAGAGATAGTTTCGGACGTTATGAGG
 ATACAGAGCGTATCAAACGTGCTGTTGAGACAAAGTCTGGCGGCAGGAATTTAACGAGAG
 ATATAGGAGGTCAGGCTTCAACAAAGGAAATGATGGAAGCTATTATTGCAAGGTTATGAA
 GTTAGACGAAAAAATTCGAT

ORF Predictions:

ORF #	Start	End	Direction	Length
1	397	486	F	30 aa

>[SEQ ID NO:129] 3860890-2 ORF translation from 397-486, direction F
 VERIIRKAFAIELQEIAEKSLLVSISKMF*

Description:

unknown

Assembly ID: 3860952

Assembly Length: 874bp

>[SEQ ID NO:40] 3860952 Strep Assembly -- Assembly id#3860952
 TCGATCTAGAGAATTGCTCCAGAGCTTCCTGACCGTCCGCTGCCTCAATAGTTTCATAGC
 CACAATCCGTCAAATAATCACTGACCCCTCACGGATCATCTCTTCATCTTCTACAATTA
 AAATTTTCATACTTTAACTGCTCTCTATTTTTTATTTTTCTTAGAATAAATACCTACTCT
 ATTTTCTATTATAGTCTCTTGCTGGCCTTTTGTATGTAAGCAACTGACCACTAGATAAAA
 CGTTGTGAAATTCCTTTCTCATAAATTCATAACTTTAGTATATTATATTAAAGCACTAA
 AGTACAAAGAAAGCAACTGAAAGCAATGATTTTCACCACTGCTTTCAGATTTATTTTGAA
 TTGTTAAATAGCTATTCCCTATCCACTATTCTTGAATAGAAACACAAGATGCAATCTTTAT
 TCCAGACTCATTTTTTAAAAAATCAAATTTATTACCATCCAGCAAGAGCTCTTTTGGTT
 GTTTTCTAAGGAGATTGCTTGAAGCAAGCGCCATAACGAGAACCACTAGAACCAAGGCAA
 GGACAAAAATGATGATAAAGTCTGATGTCTGAATGGAAATGTCTAGGCTCGACAAGGTCT
 TGCTAAAGCCATCTACTTCTGCACCGCCACCAAGGTTAGAGGCTTGAGCCGCCCTTACTAG
 CCTGTTTGGCAACACCTGAAGTCACATTGGCAAGGACAGTGTTTCCAATTCGCACGGGCA
 GTGTAATTAGCTAGGAAGTAAGCANAACTAGAGCAGGGATAGCAATCAAGATAGATTGCG
 GTGATGAATTGACCCAAGATACTTGCTGCTTGAGACCAATAGAGAGGAGGATTCCCCT
 TCCTTGCCGACGGGCATTGATCCAAAGACTGAGC

ORF Predictions:

ORF #	Start	End	Direction	Length
				66

 1 449 715 R 89 aa

>[SEQ ID NO:130] 3860952-1 ORF translation from 449-715, direction R
 VRIGNTVLANVTSGVAKQASKAAQASNLGGGAEVDGFSKTLSSLDISIQTSDFIIFVLA
 LVLVVLVLMALASSNLLRKQPKELLLDGE*

Description:
 unknown

Assembly ID: 3860962
 Assembly Length: 762bp

>[SEQ ID NO:41] 3860962 Strep Assembly -- Assembly id#3860962
 CTTGTAAACGGTCATAAAGTTTCTGCAAACCTACCATCCTTGCTCCATTTAGTAACCAAGTT
 ATCAAGATAGTCGTTGAGCTCTGTATTTGATTTCTTGGTAAACAATACCGTAGTCAGATGG
 CTTGAAACTATCATCTAGTAGTTCTGTGCGTTTAACTAGTGTAGCCAGATAGAATAGAGC
 GGTCAACGGAAAAGGCATCGATACGATGAGCGTGAAGGGAAGTAATCAATTCTGGGTAGG
 AACCAAGTTCGACGAATTTAAACTTCAGACCTTTCTTTTTACCCAGTTCAGTAATCAGGC
 GTTGGGTGATAGAACCTTGGGCGACTCCGATGGTTTTGCCGTTTAGGTCCTCAATCTTTT
 TGATTTTGGCAGATTTATTGACCAAAAATCCAGAAGCGTCTGTGTAGTAGGGACTGGTAA
 AGTTGTAGAGTTTTTTGCGTTCGTCCTGATGGTAAAGGTCGCGATATCCATATCGACCT
 GTTCATTGTCTAGAAGGGGGCCGCGGGTTTGTGCTGTAACCGGCACATAGTGAATCTTGA
 CCTTGAGTTCATCAGCTACCATTTTGGCCAAGTCGGTTTCGATACCAGAATAAGTACCGG
 TCTTGGGATCTTTGTTAACCAAAATTGGGAACGTCTTGTGTTTGACACCCGACAACCAGTTC
 GCCTCTTTTTTTGAATGTCTGCGATACTAGTATTAGCCTGGACTGGTTTGGCAGCAACAAG
 GCCGAAAAGGCTAATCAATAATGCTGATAAAAAGAATTCGAT

ORF Predictions:

ORF #	Start	End	Direction	Length
1	152	646	R	165 aa

>[SEQ ID NO:131] 3860962-1 ORF translation from 152-646, direction R
 VSNKTFPILVNKDPKTGTYSGIETDLAKMVADELKVKIHYVPVTAQTRGPLLDNEQVMD
 IATFTITDERKKLYNFTSPYYTDASGFLVNKSAKIKKIEDLNGKTIGVAQGSITQRLITE
 LGKKKGLKFKFVELGSYPELITSLHAHRIDAFSVDRSILSGYTS*

Description:
 cell adhesion factor PEB1 precursor - Campylobacter jejuni

Assembly ID: 3861268
 Assembly Length: 1942bp

>[SEQ ID NO:42] 3861268 Strep Assembly -- Assembly id#3861268
 CTCGAATTTTGGTGCTCCAGAAACGGTTCCAGCAGGAAGCGTTGCTTTCAAGGCATCCA
 TGGCAGTGAGTTCTGCAAGCAAACGTCCCTTGACCACACTGGTCAAATGCATGACGTAGC
 GGAAGAGCTCCACCTCCATATACTTAGTAAGTTGGACACTGGCCGTTTCAGAGATGCGGC
 CAATATCGTTACGCCCCAAGTCTACCAACATTCGATGTTCTGCTGTTTCCTTCTCATCAG
 AGAGGAGGTCAGTCGCCAAGGCCTTGTCTTCTCCATCCGTAGCCCCCTCTTGGTCGCGTCC
 CTGCAATCGGATTGGTTGTCACGATGCCATTTTGGACAGAAACCAAACCTTCTGGACTAG
 CTCCGATGATTGATAATCCCCAAAATCATACAAATAAAGGTAATTAGATGGATTAGTCA
 CGCGGAGATTTCTGTAGAAGTCAAATGGATTTCAGTTAACTTCTGCGTGAAGAAAACGC
 TGGCTGAGTTACACATCGGAACATATCTCCGTTACGAATCAAGTCACGAGCTGTTTCTAC
 CATTCCTCAAACCTTATGTGGAGCGATATGCGGTTTGAAGTCAAGTGGTGATAAATCCAA
 GTCTTCAAATTCATTTGGAGCAGGAATGCGTAATTCCTCAAGCACTGGTTCAAGGATTT
 TTCCAAGGCCTCTTGACTGCGCTCACTATAAAGTGCATCCTCTATGACATGTTATCTTCT
 CCTTCTTGTTGGTCAAAGACCATATAGCTCTCATAGACAAAGAAATGCATGTGCGGCGTC
 CCAATTGTATCCTCAGGGATTTGACCAATTTCTTCATAAAGCGAAATCATATCGTAACCA
 ACAAACCAATGGCTCCCCACCAAAGGGAGGTCTGAATGGTGCTGGCTCTTATGAATC
 ACTTCATAAAGGAAATCCAAGGGATCCCGATCAATCGCTTGACCATTTTGATAGAGAACT
 CCATTTTCAAACCTTAATCTCAAAACTGGATTATAGGCTAGGATAGAAAACGAGCTGTT
 TCCTTGTCTCTCGGAATACTCTCTAAAATAACCTTATGTTGCCCTTTAAGCGCATATAA
 GCCAAGATTGGTGATAAGACATCTCCATGAATGATTTCGTTCCATTGTCATTTCCCTTTCA
 GTTCTAATTCGAGTTCGTGGCGACTGTATGAAAAATCCCCACGCAAAATAACTTGCGTGA
 GGACGAAATTGCGGGTGCCACCTCAATTATAGGATTTCTCCTATCTCTCATTCCTGTCTC
 AGATATCTCCTGTAACAGGCTGTGCGATAAAGGGCACTCCCTTGAGAATGATGTTTTCTT
 CTCTCGTTTTAGATGAACCCAACTTTACAGCTTTCTCTGCTTGTTTTAGCAACCACAAG
 CTCTCTGTGAGAGAAAAGACTGTAATTTTTCCATCTATTATTTTTTAGCTTCTAGTAATC
 TGCAATCGCAGCTAGGTCCTTGCCCTCCACGACCAGAGACATTGATGAAGAGATGTTTCATC
 TCGGTACACCTTTATACTCTTCGAAAATCTCTTCAAACCGCGTCAACGTCGCCTTGCCGT
 AGGTATGGTTACTGACTTCGTTCAGTTCTATCTGCAACCTCAAACAGTGTTTTGAGCTGA
 CTTTCGTACAGTCTTATCGACAACCTCAAACAGTGTTTTGAGCAGCCTGCAGCTAGTTTCC
 TAGTTTGCTCTTTGATTTTCATTGAGTATTATTTTCATTTTCTCCTGCAATTGAATTCCTG
 CTCAGCTTTTTGTCTTCTATTTCTTTAAATCAAAGTAGCTCTTTTGTTAATAACTCGAT
 CAACAAACATCGTGGTACAAGTATCTACTTTGAAATTTATCAACCACTTAACAACGATA
 CTGTATTTCTAGGAAAACGATGACATTCTTCCTAATAAACTTCTCATATATAGCATAAA
 TTTCTACTCTTTTAAATTCGAT

ORF Predictions:

ORF #	Start	End	Direction	Length
1	457	645	R	63 aa

>[SEQ ID NO:132] 3861268-1 ORF translation from 457-645, direction R
 VLEELRIPAPNEFEDLDLSPLDFKPHIAPHKFEGMVETARDLIRNGDMFRCVTQPAFSSR

RS*

Description:

ANTHRANILATE SYNTHASE COMPONENT I (EC 4.1.3.27). - LACTOCOCCUS LACTIS
(SUBSP. L ACTIS) (STREPTOCOCCUS LACTIS).

Assembly ID: 3861270

Assembly Length: 1048bp

```
>[SEQ ID NO:43] 3861270 Strep Assembly -- Assembly id#3861270
CTGTTAAGATTGTTTCCGTGCATCCACATAGGATTTACCTTGTCTGTATGGGCCAATTCA
CCCATCAAAAACGCCATAGGTCTCATCTGTCAAGATACTAGACATACCGATATTGTACCAA
AGACTGGTATGACGGAAATAAGTCGATGCGTGTAACACTCAACAAAAGAGACGCAAGTTG
ATTAGAAAAACCGTCATAGCAATAGCTGCCACAGGAGCTTGAACCACAATCAGTGCCAAC
ATGGCAAACACTGGGCACTCCCAGCATAAACAAGAGACTCATCAAGCCCATCTCAACAGGT
GTCACATAGGGCGCACCGATAGTCCCACAGGCCAGGCCGATACTGACATAGCCAAGAGCC
GTTGGCATGGCTGCCTGCGCCCCCTCCTAAAATCCTTTTCTTTTCATCTTTCTCCTCATA
TTGTCTTAATAATACTCAATGAAAATCAAAGAGCAAAC TAGGAAATTAGCCGCAGGNTGC
TCAAAACACCGTTTTTGAGGTTGCAGATAGAACTGACGAAGTCAGCTCAAAACACCGTTT
TGAGGTTGCAGATAGAACTGACGAAGTCAGTAACATATATACGGCAAGGCGACGTTGACG
TGGTTTGAAGAGATTTTCGAAGAGTATTAGAAAATGCCGATAAGGGTCTGCATACCAAGG
CTGGTGAGGATGATGGCAATCCAGCAGACGGCTCCGAGAACAAATGGATTTTCCACTGGAT
TTGACCATAGCGACCAGATTAGTTTTGAGACCGATGGCACTCATGGCCATGATAATGAGG
AATTTAGAGAGTTGTTTGAGAGGGGTAAAGAACTACTAGACACACCGAGAGAGGTCAGA
AGGGTGTTAGGAGCGATGCAAGGATGAAGTAAAGGATAAAAAGTGGGAAGACTTTTTTC
AGTTGTAAGCCTTGCTTATTTTTTTTGCTCGCGACTTTGCCAGTAGGAGAGAAAGAGAGTG
ATGGGGATGATAGCTAGGGTGCGCGTGAGTTTGACAATGGTTGCGGATTCGAGGGTATTG
GTCTGGTAGAGACTGTCCCAAGCGCTAG
```

ORF Predictions:

ORF #	Start	End	Direction	Length
1	627	824	R	66 aa

```
>[SEQ ID NO:133] 3861270-1 ORF translation from 627-824, direction R
VSSSFFTPLKQLSKFLIIMAMSAIGLKTNLVAMVKSSGKSIVLGAVCWIAIILTSLGMQT
LIGIF*
```

Description:

unknown

Assembly ID: 3861288

Assembly Length: 1571bp

>[SEQ ID NO:44] 3861288 Strep Assembly -- Assembly id#3861288
 AGAGCTGGTAATATTCCCAAAGAAACGGCTCAAATCGAATTAGAAAGCCTTCTGCAAAAA
 GGAATCCCAGTCGCTCTGGTATCACGATGCTTTAACGGTATTGCCGAGCCTGTTTATGCC
 TACCAGGGTGGGGGCGTACAGTTGCAAAAAGCAGGCGTTTTCTTTGTTAAAGAACTCAAC
 GCCCAAAAAGCCCGCTTGAAACTCCTCATCGCCCTCAATGCCGGACTAACAGGACAGGCT
 TTGAAAGACTATATGGAAGGCTAATACTCTTCGAAAATCTCTGCAAACCACGTCAGCGTC
 GCCTTACCGTATGTAGAGCACAAAATCAGGAAATCTTCTCGATTCCCTGATTTTTTCTAT
 TTACGTTTTCTGTGTTGAGCTACGTTCTGTCAAACCATGAGGTAAGAGAACTTCACGTTCT
 TCCAACTCTTCCTTATGCATAATCTTGGTCAACATACGCATACTAATGGCACCAAGGTCA
 TAAAGAGGTTGGGCAATCGTTGTCAAGTTTGGACGGGTAAAGCGTGAGATTTGTGAATCA
 TCACTAGTAATAATTCGATAATCTTCTGGCACAGAAACACCTTATCAGCCAAACCGTTCA
 AGACTCCTGCTGCCAACTCATCACCTGTCACAACTGCTGCAGTTGCATTTGATGAAATCA
 AACGCTCTGCTAAGGCGTAACCATCATCATAGCTATATTTAGATTCAAATACCAAACCTT
 CACTATAAGCGATTCCCTGCTTTTTTCAAGGTTTCTTGTAGCCAACTAAACGAACCTTAC
 CATTGATGTCATCCACTAGCGGACCGCTAACGAAAGCAATACGCTCATTTTCTTTAGCAA
 GGTAACCTCACTGCATCAATTGTTGCTTGCTTATAGTCAATATTGACACTTGGCAACTGGT
 GCTCAACATCGACAGTTCTCTGCGAGAACAATCGGAGTACGTGAACGCGAAAATTCTGAGC
 GAATTTTATCTGTCAAGTGATAACCCATATAGATAATGCCATCTACCTGCTTTGAAAAGA
 GGGTATTGACAACAGAACTTCTTTCTCGTTATCTTCATCGCTATTAGCTAGGACAATAT
 TGTACTTGTACATTTCTGCAATATCATCAATCCCTTAGCCAAACTCGAAAAATAACCAT
 TGGTAATATTTGGAATCACGACACCGACAGTGGTTGTCTTTTTTACTTGCAAGACCACGCG
 CAACTGCATTTGGACGATAATCCAAACGATCAATTACCTCTAGCACTTTTTTACGGGTAT
 TCTCTTTTACATTTTATTGCCATTGACCACACGGCTGACCGTCGCCATGGGAAACACCT
 GCTTCACGAGCGACATCATAAATGGTTACTGTATCATCTGCATTCATTCCTTTTCTGTC
 CTTTCTATCTCCACACATTCTTTTACAAGTAGAAGTGCTGAATTGAAAGCTCTATATCTT
 ACTTACAAAATGAAGATGTGAAAATTTCTGTTTTTCATATTTCTACTTATTCCATTCTATC
 ACTAATTGTAAACACTTTCAAGTGTTTTTTGAAGATTGATTGAAAAAATTTCATAGAAAA
 CCTAGGTTTAG

ORF Predictions:

ORF #	Start	End	Direction	Length
1	357	572	R	72 aa

>[SEQ ID NO:134] 3861288-1 ORF translation from 357-572, direction R
 VPEDYRIITSDDSQISRFTRPNLTTIAQPLYDLGAISMRMLTKIMHKEELEEREVLLPHG
 LTERSSTRKRK*

Description:

GLUCOSE-RESISTANCE AMYLASE REGULATOR. - BACILLUS SUBTILIS.

Assembly ID: 3861306

Assembly Length: 1682bp

```
>[SEQ ID NO:45] 3861306 Strep Assembly -- Assembly id#3861306
CTGACGTAAAAAAGATTTTTCGGAAAAGTATCATCATCTATTTTAGACCATTTTCTTATAA
TAACCATTTTATTTTATTTTGTCAAGGTCTTTGAATTCTTTCTTAAACAAGCCTTGTAAT
CTCTACTTTTGAAGAATTTATTTTTCCTTACTGACAAGATTGAGACGGTAGGAATCATT
GAAAATAACCTAGCCAACATCAATCACAATCATTCTCCTTTCTCAATTACACTAAATTA
TAGTGTATTGAATCTATAACAGTGCACCTTGGCTGCTAAAATATTTCTATAAATTAATTT
GACTTTCCTGATAGAGTTGTTTACATCTTATTTCAATTCACTATACTTTCCCTTATACTC
AATGAAAATCAAAGCGCAAAGTAGGAAGCTAGCCACAGGCTGCTCAAAGCACTGCTTTGA
GGTTGTAGATAAGACTGACGAAGTCAGTTACATATATCTACGGCAAGGCGAAGCTGACGC
GGTTTGAAGAGATTTTCGAAGAGTATAAAGTTTGGTTTCTGTATCTTTCAGAAAAATAAGG
TATACTGTATGTAAACGATTTCAAAGGAGTCCAGTTATGGCAAAAACATTTTTTTATTCCA
AATAAACAGAGCATTTTAGGAGAACAAGAGATTTTGAATGCCAAGTCGATCTTGCTATG
ATGTAGTCTATCTCCGTCAGCCTCTTAATCGTCTCGAGTATATTGAGTGTGCGATAGTGG
GGCAATCACAATTTCTTTTTAAGGTCAGTTATGCTGATGGTCAAAAGGCTTACCGTGTCTG
ATCTTCCTGACCTACTAACAAGACAGACTGGCAGATTATCAAGTCATTTTTTAGATGTTT
TGCTTGCTTATACAGGGACTGATATTGAAGGGCTAGATGGTTTTGATTTTGAAGCTTATT
TCCAAGCAAGTATTCAAGCCTATCTAGCAGACCCTGTAGCTCGTTTTACGATTTGCCAAC
GAATTTTTAATCCTATTTTCTTTAGTCGTGAGAACTTGAAAAGCTTTTAGAGGCAGATG
GCTTGCTCAGTTTGAAGCGCGTGTGCGTGCGGTTCAAGAGACAGATGCCTACTTTGCGA
GAGTTTCCTTCTATCAGGATGGAGAAGGAAAAGTGCATGGCGTTTACCATCTAGCTCAAG
GAGTCAAGACAGTTTACCGAGAGAACCGTTTGTTCCTGCAGCCTATATTGAGCGAATTG
GTGGATAAGGAAGTCCAGTGGGAGATTGACTTGGTTCAAATCACAGGAGACGGCTCTAAA
CCAGAAGACTATGAATCCATAGCTCGCTTGGACTATGCAAAATCTTAGAGGTATTACCC
CCATCTTTTTTACCACCAACTAGACGCCAATCAAATAGAAATACAACCCATCCTAGGACAA
GATTTTAAAACATTAGCACAAAGAAAAGTAAAGCAGAAGCAGGTCAATCGACTTGCTTTTT
TGACATAGAAAAATCCTGCCAAGGATGACAGGATTGCTACTCAATGAAAATCAAAGAGC
AACTAGGAAGCTAGCCGCAGGCTGTACTTGAGTACGGTAAGGCGAAGCTGACGTGGTTT
GAATTTGATTTTCGAAGAGTATGAATTTTAAAGAAAGGCCAAGATACGAAGATAATCTCC
AATCAGTGCCACTTCAGCTTCCAAGAAGAAGAAGATTATAACTCCCGTTCCCAAGGACA
GA
```

ORF Predictions:

ORF #	Start	End	Direction	Length
1	717	1208	F	164 aa
2	1201	1410	F	70 aa

```
>[SEQ ID NO:135] 3861306-1 ORF translation from 717-1208, direction F
VGQSQFLFKVSYADGQKAYRVDLPDLLTKTDWQIIKSFLDVLLAYTGTDIEGLDGFDFEA
YFQASIQAYLADPVARFTICQIRIFNPIFFSRENLSKFLEADGLAQFEARVRAVQETDAYF
ARVSFYQDGEKGVHGVYHLAQGVKTVLPREPFVPAAYIERIGG*
```

Description:

unknown

>[SEQ ID NO:136] 3861306-2 ORF translation from 1201-1410, direction F
VDKEVQWEIDL VQITGDGSKPEDYESIARLDYAKFLEVLPPSFYHQLDANQIEIQPILGQ
DFKTLAQEK*

Description:

unknown

Assembly ID: 3861334

Assembly Length: 3041bp

>[SEQ ID NO:46] 3861334 Strep Assembly -- Assembly id#3861334
ATCGAATTAAAAATGAGGTATTCAGGCTTGTGATTTTCTATGGAAGTTAATAGTGATTGC
CTCTAATGCTTACAAGTGATATTA AAAATAGAGGACCTAGTGATGTCAATCATTTCAACT
GATTTAACCCCTTTTCAAATAGATGATACATTGAAAGCAGCCTTGCGAGAAGATGTTTCAT
TCCGAAGATTACAGTACCAATGCCATTTTGTATCATCATGGCCAAGCCAAGGTGTCGCTT
TTTGCCAAGGAAGCTGGTGTTTTAGCGGGGCTAACCGTTTTTCAAAGGGTTTTTACCCTA
TTTGATGCCGAGGTGACCTTCCAGAATCCTCATCAATTTAAGGATGGGGATCGTTTGACT
AGTGGCGATTTGGTTTTAGAAATCATAGGCTCGGTGAGAAGTCTCTTAACATGTGAACGC
GTTGCCTTGAATTTTTTACAACATTTATCAGGGATCGCTTCGATGACAGCTGCTTATGTA
GAAGCCTTAGGCGATGATTGCATTAAGGTATTTGATACTCGAAAACTACTCCTAATTTA
CGTCTTTTTGAGAAATATGCCGTGAGAGTTGGCGGTGGCTATAATCATCGCTTTAATTTA
TCAGATGCTATCCTGCTAAAAGACAATCACATTGCGGCAGTAGGTAGTGTCAAAGGGCA
ATTGCTCAAGCGCGTGCCATGCTCCTTTTGTGAAAATGGTCGAGGTGGAAGTGGAAGC
CTTGCTGCTGCCGAAGAAGCTGCGGCGGCGGGTGCTGATTTATCATGTTGGATAATATG
TCATTGGAACAGATTGAACAGGCCATTACCCTAATTGCAGGACGTTCTCGGATTGAATGT
TCTGGAAATATTGATATGACCACTATTAGCCGTTTTCGTGGTTTAGCGATTGATTACGTC
TCCAGTGGTAGTTTAACCCATAGTGCTAAGAGTCTTGATTTTCCATGAAGGGTTTAACC
TACCTTGATGTCTAAGTTGTAAAATAAACTAACTTTTTAAAGGATGTCTTTCCTCTAGAA
CGAGTTTTATGTGATAGATAGTTTAAACGCCTCTTCAAATATAGTAAAATGAACCAAAAATA
GTACACAATGTGGTATAATCTTCTTATGGCATATTCAATAGATTTTCGTAAAAAAGTTCT
TTCTTATTGTGAGCGAACAGGTAGTATAACAGAAGCATCACACGTTTTCCAAATCTCACG
TAATACCATTATTTGAGCTGGTTAAAGCTAAAAGAGAAAACAGGAGAGCTAAACCACCAAGT
AAAAGGAACAAAACCAAGAAAAGTTGATAGAGATAGACTTAAAACTATCTTACTGACAA
TCCAGACGCTTATTTGACTGAAATAGCTTCTGAATTTGGCTGTCATCCAACCTACCATCCA
CTATGCGCTCAAAGCTATGGGCTACACTCGAAAAAAGGACCACACCTACTATGAACAAGA
CCCAGAAAAAGTAGCCTTATTTCTTAAAAATTTTAATAGTTTAAAGCACCTAGCACCTGT
TTAGATTGATGAAACAGGATTCGATACTTATTTTTATCGAGAATATGGTCGCTCATTA
AGGTCAGTTAATAAGAGGTAAAGTATCTGGAAGAAGATATCAGAGGATTTCTTTGGTTGC
AGGTCTAACAAATGGTGAGTTAATCGCTCCAATGACTTACGAAGAGACGATGACGAGCGA

CTTTTTTGAAGCATGGTTTCAGAAAGTTTCTCTTACCAACATTAACCACACCATCGGTTAT
TATTATGGATAATGCAAGATTCCATAGAATGGGTAAGTTAGAACTTTTATGCGAGGAGTT
TGGGCATAAACTTTTACCTCTTCCCTCCCTACTCGCCTGAGTACAATCTTATTGAGAAAAC
ATGGGCTCATATCAAAAAGCACCTCAAAAAGGTATTACCAAGTTGCAATACCTTTTATGA
GGCTCTTTTGTCTCTGCTCTTGTTCATTGACTATAGTTCACGGATACAGTTGGGAAAGA
AGTTAAATGTAGTTGGATTTCCTACTAAAGGTTGATGAGTAAGTTTTTGTATCTGAACCTG
ATTGGCCGCAAGCAGCTAAAAGCAAAGCAGATGCAAAAGTCAGACCTGCACCAAGGACAC
GCTTCTTTATGTTTCATCTTCTTTCTCCTTAATAGTGGGAATTTGTAAAGTTAATTGAATT
TCAAGAATGAAGTTTTATAAACTTTGGTTATAAAAAACAAAGGATTTCTGTCTTTTATA
CAGTCCTCCCCTTGTTTTTATACGATTTCAATTTTAAATTTTTCTGCAAAAAATATTTAT
AGTAATTCCACACAGAAAGCATCCCATGGAACCTAAGATTTGTTTTTCAAAGACTTCTTGA
GCTAGGGTGTTTTCAATCAAGACAGATTTGACTTTTCCCTTCTACTGTCAAGTCTTGCTCT
TCATTGGACAAGTTAGCCACAACCTAGGAAGCGACGGTCGCCATCCTTACGTATATAAGCA
AAGACCTTATCAGCCGTATCAAGCAATTCAAAGTCAGCTCGAATTAGCCAACCTATTCTCC
TTGCGAATTTGGACCAGTTTCTGATAGGTATAGAAAATAGAATCTGGATTTGCCAGCGCT
TCTTGGACGTTGATCATCTCGTAATTTGGATTAACTGCCAACCAAGGTTGACCTGTTGAG
AAACCAGCGTTTTTGTCTCTCGTCCCATTGCATAGGGGTACGGGCATTGTACAGTCCAATA
ACACGGATACTGTCCATGATTTCTTGCATCGGAACACCTTTTTTCAAGAGCCTCACGCGCA
TAGTTGAGAGATTCAATATCTTCTACTTGATCCAGTGTTTCAAACGGATAGTTGGTCATC
CCAATCTCCTCACCTTGGTAGATATAAGGAGTTCCTCTCATAAGATGAAGCAAGATTGCA
AAGGCTTTGGCAGATTTTTTCGCGGTATTCTTGGTCATTTCCCCAGATTGAGACAATACGA
GGGAGGTCATGGTTGTTCCAGAAGAGGGAATTCCAGCCGTCCTCAACTCCTAACTCTGTC
TGCCATTTGTTGAAGATTTCTTTTAACTTAGCGATATTTCAG

ORF Predictions:

ORF #	Start	End	Direction	Length
1	76	975	F	300 aa

>[SEQ ID NO:137] 3861334-1 ORF translation from 76-975, direction F
VILKIEDLVMSIISTDLTPFQIDDTLKAALREDVHSEDYSTNAIFDHHGQAKVSLFAKEA
GVLAGLTVFQRVFTLFDAEVTQNPQHFKDGDRLTSGDLVLEIIGSVRSLTLCERVALNF
LQHLSGIASMTAAYVEALGDDCIKVFDRKTTPNLRLFEKYAVRVGGGYNHRFNLSAIL
LKDNHIAAVGVSQRAIAQARAYAPFVKMVEVEVESLAAAEAAAAGADIIMLDNMSLEQI
EQAITLIAGRSRIECSGNIDMTTISRFRGLAIDYVSSGSLTHSAKSLDFSMKGLTYLDV*

Description:

PROBABLE NICOTINATE-NUCLEOTIDE PYROPHOSPHORYLASE (CARBOXYLATING) (EC
2.4.2.19) (QUINOLINATE PHOSPHORIBOSYLTRANSFERASE (DECARBOXYLATING))
(QAPRTASE) (FRAGMENT). - BACILLUS SUBTILIS (BLAST)

Assembly ID: 3864148

Assembly Length: 4694bp

>[SEQ ID NO:47] 3864148 Strep Assembly -- Assembly id#3864148
TTAATTTAAATTCTTAAATTTTTTTCATAATAATCTCCCTATAAAAAATAAGTCGCCCAA
TCAGGCGGCTTATTTTTTTTGAAAAATGGGCTTGGTGCCTGAGAATAAATAGCTTAGTGAT
AGAAGAAAATGGGGAAATATGGTATAATGAAACGATAGATTTTTTGAATAGGAATAAGATC
ATGTTTGGATTTTTTAAGAAAGATAAAGGCTGTGGAAGTAGAGGTTCCGACACAGGTTCC
TGCTCATATCGGCATCATCATGGATGGCAATGGCCGTTGGGCTAAAAACGTATGCAACC
GCGAGTTTTTGGACATAAGGCGGGCATGGAAGCATTGCAAACCGTGACCAAGGCAGCCAA
CAAACCTGGGCGTCAAGGTTATTACGGTCTATGCTTTTTTCTACGGAAAACCTGGACCCGTCC
AGATCAGGAAGTCAAGTTTATCATGAACCTGCCAGTAGAGTTTTATGATAATTATGTCCC
GGAAC TACATGCGAATAATGTTAAGATTCAAATGATTGGGGAGACAGACCGCCTGCCTAA
GCAAACCTTCGAAGCTTTAACCAAGGCTGAGGAATTGACTAAGAACAACACAGGATTGAT
TCTTAATTTTGCTCTTAACTATGGTGGACGTGCTGAGATTACACAGGCGCTTAAGTTGAT
TTCCCAGGATGTTTTAGATGCCAAAATCAACCCAGGTGACATCACAGAGGAATTGATTGG
TAACTATCTCTTTACCCAGCATTTGCCTAAGGACTTACGAGACCCAGACTTGATTATCCG
TACTAGTGGAGAATTGCGTTTGAGCAATTTCTTCCATGGCAGGGAGCCTATAGTGAGCT
TTATTTTACGGACACCTTATGGCCTGATTTTGACGAAGCGGCCTTGCAGGAAGCTATTCT
TGCCCTATAATCGTCGCCATCGCCGATTTGGAGGAGTTTAGGAGGAAATATGACCCAGGAT
TTACAGAAAAGAACCTTGTTATGCAGGGATTGCCCTGACTATTTTCTACCAATTTTAAT
GATTGGGGGCTCTTGCTTCAGATAGCAATCGGAATCATANCCATGCTAGCCATGCATGAA
CTTTTGAAGATGAGAGGTCTAGAGACCATGACGATGGAGGCCTCTTGACCCTCTTTGCAC
NTTNGTATTGACCATTCCCCCTGGAATCGAATTACCTGACTTTTTTGCCAGTTGATGGGAA
TGTGGTTGCCTATAGTGTTTTGATTTCAATCATGTTAGGAACGACCGTTTTTAGCAAGTC
TTATACGATTGAGGATGCGGTTTTCCCTCTTGCTATGAGCTTCTACGTGGGCTTTGGATT
TAATGCTTTACTAGATGCTCGTGTTGCAGGTTTGGACAAGGCTCTCTTAGCCTTGTTGTAT
CGTCTGGGCGACAGACAGTGGTGCCTATCTTGTTGGGATGAACTATGGGAAACGAAAGTT
AGCACCAAGGGTATCGCCTAATAAAACCCTTGAGGGTGCCTTGGGTGGTATTTTAGGAGC
AATTTTAGTAACCATTATCTTTATGATAGTTGACAGTACAGTTGCTCTTCCATATGGAAT
TTACAAGATGTCAGTCTTTGCTATTTTCTTTAGCATTGCTGGACAATTTGGTGATTTACT
AGAAAGTTCGATCAAACGTCATTTTGGTGTTAAGGATTCTGGGAAATTTATCCCTGGACA
TGGTGGTGTTTTGGATCGTTTCGATAGTATGTTGCTTGTATTTCCAATCATGCACTTATT
TGGACTCTTTTAATCAAAAGACGGAGGAAACGCTATGCTCGGAATTTTAACCTTTATTCT
GGTTTTTGGGATTATTGTAGTGGTGCACGAGTTCGGGCACTTCTACTTTGCCAAGAAATC
AGGGATTTTAGTACGTGAATTTGCCATCGGTATGGGACCTAAAATCTTTGCTCACATTGG
CAAGGATGGAACGGCCTATACCATTCGAATCTTGCCCTCTGGGTGGCTATGTCCGCATGGC
CGGTTGGGGTGATGATACAACTGAAATCAAGACAGGAACGCCTGTTAGTTTGACACTTGC
TGATGATGGTAAGGTTAAACGCATCAATCTCTCAGGTAAAAAATTGGATCAAACAGCCCT
CCCTATGCAGGTGACCCAGTTTGATTTTGAAGACAAGCTCTTTATCAAAGGATTGGTTCT
GGAAGAAGAAAAACATTTGCAGTGGATCACGATGCAACGGTTGTGGAAGCAGATGGTAC
TGAGGTTCCGATTGCACCTTTAGATGTTCAATATCAAATGCGACTTTATCTGGGGCAAA
CTGATTACCAATTTTGCAGGTCCATGAACAATTTTATCTTAGGTGTTGTTGTTTTTTGG
GTTTTAATCTTTATGCAGGGTGGTGTGAGAGATGTTGATACCAATCAGTTCCATATCATG
CCCCAAGGTGCCTTGGCCAAGGTAGGAGTACCAGAAACGGCACAAATTACCAAGATTGGC

TCACATGAGGTTAGCAACTGGGAAAGCTTGATCCAAGCTGTGGAAACAGAAACCAAAGAT
AAGACGGCACCGACTTTGGATGTGACTATTTCTGAAAAGGGGAGTGACAAACAAGTCACT
GTTACACCCGAAGATAGTCAAGGTCGTTACCTTCTAGGTGTTCAACCGGGGGTTAAGTCA
GATTTTCTATCCATGTTTGTAGGTGGTTTTACAACCTGCTGCTGACTCAGCTCTCCGAATT
CTCTCAGCTCTGAAAAATCTGATTTTCCAACCGGATTTGAACAAGTTGGGTGGACCTGTT
GCTATCTTTAAGGCAAGTAGTGATGCTGCTAAAAATGGAATTGAGAATATTCTTGTACTT
CTTGGCAATGATTTCCATCAATATTGGGATTTTTAATCTTATTCCGATTCCAGCCTTGGA
TGGTGGTAAGATTGTGCTCAATATCCTAGAAGCCATCCGCCGCAAACCATTGAAACAAGA
AATTGAAACCTATGTCACCTTGGCCGGAGTGGTCATCATGGTTGTCTTGATGATTGCTGT
GACTTGGAATGACATTATGCGACTCTTTTTTAGATAATCGAGGAATATTATGAAACAAAG
TAAATGCCTATCCCAACGCTTCGCGAAATGCCAAGCGATGCTCAAGTTATCAGCCATGC
TCTTATGTTGCGTGCTGGTTATGTTGCGCAAGTTTCAGCAGGTGTTTATTCTTATCTACC
ACTTGCCAACCGTGTGATTGAAAAAGCTAAAAACATCATGCGCCAAGAATTCGAAAAGAT
TGGTGCTGTTGAGATGTTGGCTCCAGCCCTTCTTAGTGCAAGATTGTGGCGTGAATCAGG
TCGTTACGAAACCTATGGTGAAGACCTTTACAACCTGAAAAACCGTGAAAAATCAGACTT
TATCTTAGGTCCAACCTCACGAAGAAACCTTTACAGCTATTGTCCGTGATTCTGTTAAATC
TTACAAGCAATTGCCACTCAACCTTTATCAAATTCAGCCCAAGTATCGTGATGAAAAACG
CCCACGTAATGGACTTCTTCGTACACGTGAGTTTATCATGAAGGATGCTTATAGTTTCCA
CGCTAACTATGATAGTTTGGATAGTGTTTATGATGAGTACAAAGCAGCCTATGAGCGTAT
TTTCACTCGTAGTGGTTTAGACTTCAAGGCTATTATTGGTGACGGTGGAGCCATGGGTGG
TAAGGATAGCCAAGAATTTATGGCCATTACATCTGCTCGTACAGACCTTGACCGCTGGGT
TGTCTTGGACAAGTCAGTTGCCTCATTGTGACGAAATTCCTGCAGAAGTGCAAGAAGAAAT
CAAGGCAGAATTGCTCAAATGGATAGTCTCTGGTGAAGATACCATTGCTTACTCAAGTGA
GTCTAGCTATGCAGCTAACTTAGAAATGGCAACAAACGAGTACAAACCAAGCAACCGTGT
TGTCGCTGAAGAAGAAGTTACTCGTGTTGAAACGCCAGATGTTAAATCAATTGATGAAGT
TGCAGCCTTCCTCAATGTTCCAGAAGAACAAACGATTAAAACCCCTCTTCTACATTGCAGA
TGGTGAGCTTGTTGCAGCCCTTCTAGTTGGAAATGACCAACTCAACGAAGTCAAGTTGAA
AAATCACTTGGGAGCAAATTTCTTTGACGTTGCTAGCGAAGAAGAAGTGGCGAATGTTGT
TCAAGCAGGATTTGGTTCACCTTGACCAGTTGGTTTGCCAGAGAATATTAATAATTATTGC
AGATCGTAAGGTGCAAGATGTTTCGCAATGCAGTTGTTCGGTGCTAACGAAGATGGCTACCA
CTTGACTGGTGTGAACCCAGGCCGTGATTTTACTGCAGAATATGTGGATATCCGTGAAGT
TCGTGAGGGTGAAATTTCCCCAGATGGACAAGGTGTCCTTAACCTTTGCGCGTGGTATTGA
GATCGGTCATATTTTCAAACCTCGGAACCTCGCTATTTCAGCAAGCATGGGAGCAGATGTCTT
GGATGAAAAATGGTCGTGCTGTGCCAATCATCATGGGATGTTACGGTATCGGTGTCAGCCG
TCTTCTTTCAGCAGTGATGGAGCAACACGCTCGCCTCTTTGTTAACAAAACGCCAAAAGG
TGAATACCGTTACGCTTGGGGAATCAATTTCCCTAAAGAATTGGCACCATTGATGTGCA
TTTGATTACTGTTAATGTCAAGGATGAAGAAGCGCAAGCCTTGACAGAAAACTTGAAGC
AAGCTTGATGGGAG

ORF Predictions:

ORF #	Start	End	Direction	Length
1	212	940	F	243 aa

2	1202	1753	F	184 aa
3	2750	3037	F	96 aa

>[SEQ ID NO:138] 3864148-1 ORF translation from 212-940, direction F
 VEVEVPTQVPAHIGIIMDGNGRWAKKRMQPRVFGHKAGMEALQTVTKAANKLGVKVITVY
 AFSTENWTRPDQEVKFIMNLPVEFYDNYVPELHANNVKIQMIGETDRLPKQTFEALTKAE
 ELTKNNTGLILNFALNYGGRAEITQALKLISQDVLDKINPGDITEELIGNYLFTQHLPK
 DLRDPDLIIRTSGELRLSNFLPWQGAYSELYFTDTLWPDFDEAALQEAILAYNRRHRRFG
 GV*

Description:

unknown

>[SEQ ID NO:139] 3864148-2 ORF translation from 1202-1753, direction F
 VVAYSVLISIMLGTTVFSKSYTIEDAVFPLAMSFYVVGFGFNALLDARVAGLDKALLALCI
 VWATDSGAYLVGMNYGKRKLAPRVSPNKTLEGALGGILGAILVTIIFMIVDSTVALPYGI
 YKMSVFAIFFSIAGQFGDLLESSIKRHFGVKDSGKFIPGHGGVLDRFDSMLLVFPIMHLF
 GLF*

Description:

CDP-diglyceride synthetase (cdsA) homolog - Haemophilus influenzae
 (strain Rd K W20)

>[SEQ ID NO:140] 3864148-10 ORF translation from 2750-3037, direction
 FVDLLLSLRQVVMMLKMEIRIFLYFLAMISINIGIFNLIPIPALDGGKIVLNILEAIRRKP
 LKQEIETYVTLAGVVIMVVLMIAVTWNDIMRLFRR*

Description:

unknown

Assembly ID: 3864172

Assembly Length: 1352bp

>[SEQ ID NO:48] 3864172 Strep Assembly -- Assembly id#3864172
 CTCGTAAGTTCGGAAGCTATCTACACAAGAAATTAACCGCTGCCTAAAGGAGAAGCCATG
 TCAACATATAACTGGGATGAGAAGCATATCCTTACCTTTCCTGAAGAAAAAGTAGCCCTT
 TCTACTAAGGATGTCCATGTTTACTATGGTAAAAATGAATCCATTAAGGGGATTGATATG
 CAATTTGAAAGAAATAAAATTACAGCTTTGATTGGTCCGTCGGGATCGGGGAAATCTACC
 TACTTACGCAGTCTCAATCGCATGAATGATACCATTGATATTGCTAAAGTAACTGGGCAG
 ATTCTCTATCGTGGAATTGATGTCAACCGTCCAGAAATCAACGTTTATGAAATGCGTAAA
 CACATTGGAATGGTTTTTCAACGCCCCAATCCATTTGCTAAATCGAATTTACCGTAATAT
 TACCTTTGCGCATGAACGTGCTGGAGTTAAGGATAAGCAAGTCCTAGATGAAATCGTAGA
 AACCTCCCTTAGTCAGGCTGCCCTTTGGGATCAGGTTAAAGACGATCTCCACAAGTCAGC

CTTGACCTTATCAGGTGGTCAGCAACAACGTCTCTGTATCGCTCGTGCCATCTCTGTAA
 GCCAGATATCCTCTTAATGGATGAGCCAGCCTCAGCCTTGGATCCGATTGCGACCATGCA
 ACTAGAAGAGACCATGTTTGAGCTCAAGAAAACTTTACCATCATCATTGTAACGCATAA
 TATGCAGCAGGCTGCTCGTGCAAGTGACTATACAGGCTTCTTTTACTTGGGTGATTGAT
 TGAGTATGACAAGACTGCAACTATTTTCCAAAATGCCAAGCTACAGTCCACCAATGACTA
 TGTATCTGGTCACTTTGGTTAGAAAGGAAACCGTATGACAGATGCGATTTTACAGGTATC
 AGACCTGTCCGTTTATTATAATAAAAAGAAGGCTTTGAATAGTGTTCCTATCTTTCCA
 ACCTAAGGAAATTACAGCCTTGATTGGTCCATCTGGATCAGGGAAGTCAACCCTCCTCAA
 GTCTCTCAACCGCATGGGAGATCTCAATCCAGAGGTGACCACAACCTGGATCCGTGGTGTA
 CAATGGTCACAACATCTACAGTCCGCGTACAGATACGGTTGAATTACGTAAGGAAATCGG
 AATGGTTTTCCAACAACCTAATCCTTTCCCTATGACTATCTATGAGAATGTTGTCTACGG
 GCTTCGTATCAATGGAATTAAGGATAAGCAGGTTCTGGATGAAGCCGTAGAAAAAGCCTT
 GCAAGGTGCCTCTATCTGGGATGAGGTCAAGGATCGTCTATATGATTACAGCTATTGGATT
 GTCAGGTGGTCAACAGCAGCGTGTCTGCGTGG

ORF Predictions:

ORF #	Start	End	Direction	Length
1	311	862	F	184 aa

>[SEQ ID NO:141] 3864172-2 ORF translation from 311-862, direction F
 VELMSTVQKSTFMKCVNTLEWFFNAPIHLLNRIYRNITFAHERAGVKDKQVLDEIVETSL
 SQAALWDQVKDDLHKSALTLSGGQQQRLCIARAI SVKPDILLMDEPASALDPIATMQLEE
 TMFELKKNFTIIIVTHNMQQAARASDYTGFFYLGDLIEYDKTATIFQNAKLQSTNDYVSG
 HFG*

Description:

HYPOTHETICAL ABC TRANSPORTER (ORF75). - BACILLUS SUBTILIS. (BLAST)

Assembly ID: 3864180

Assembly Length: 2258bp

>[SEQ ID NO:49] 3864180 Strep Assembly -- Assembly id#3864180
 AACTTCGACCGTGATAAACAAGCTGAGCTTTGACATACTTGTAGCCAACCTAAAAGCCGT
 TCTTCAAGGCCTCAAACCAGCTGCAACTCATTACAGGAAGCCTGGATGAAAATGAAGTGGC
 TGCCAATGTTGAAACCAGACCAGAACTCATCACAAGAACTGAAGAAATTCCATTTGAAGT
 TATCAAGAAAGAAAATCCTAATCCCAGCTGGTCAGGAAATATTATCACAGCAGGAGTCAA
 AGGTGAACGAACCTATTACATCTCTGTACTCACTGAAAATGGAAAAACAACAGAAACAGT
 CCTTGATAGCCAGGTAACCAAAGAAGTTATAAACCAAGTGGTTGAAGTTGGCGCTCCTGT
 AACTCACAAGGGTGATGAAAGTGGTCTTGACCAACTACTGAGGTAAAACCTAGACTGGA
 TATCCAAGAAGAAGAAATTCCATTTACCACAGTGACTCGTGAAAATCCACTCTTACTCAA
 AGGAAAAACACAAGTCATTACTAAGGGTGTCAATGGACATCGTAGCAACTTCTACTCTGT
 GAGCACTTCTGCCGATGGTAAGGAAGTGAAAACACTTGTAATAGTGTCTGCTAGCACAGGA

AGCCGTTACTCAAATAGTCGAAGTCGGAAGTATGGTAACACATGTAGGCGATGAAAACGG
 ACAAGCCGCTATTGCTGAAGAAAAACCAAACTAGAAATCCTAAGCCAACCAGCTCCTGC
 TGAGGAAAGCAAAGCTCTTCCTCAAGATCCAGCTCCTGTGGTAATAGAGAAAAAACTTCC
 TGAAACAGGAAGCTCACGATTCTGCAGGGACTAGTAGTCGCAGGACTCATGGCCACACTAG
 CAGCCTATGGACTCACTAAAAGAAAAGAAGACTAAGTCTTTTCGATAAAAAATAAACAGC
 GAGATTGAAGCTCGCTGTTTATTTTAAATTAATCACCTAGTCCAAGACGTTCAAAGATA
 TCATCCACTCGTTTGGTGTAAATAAACTGGGTTGAAGATTCATCGATTTCTTCTTGTGTG
 AGACGTGATGTTACTTCTGAATCTGCCTCAAGAAGTGGTTTAAAGTCTACTTGGTTGTCC
 CAAGAGTAGGCTGTTTTTGGTTGCACCAAGTCATAGGCTTGCTCACGGGTCATGCCTTTT
 TCAATCAATGTCAACATAGCCCGTTGGCTAAAGATAAGACCAAAAGTCGAGTTCATGTTT
 CGGATCATATTTTCTGGGAAGACTGTCAAGTTCCTTGACGATATTTCCAAAACGGTTGAGC
 ATGTAGTCAATCAAAATGGTCGTATCTGGTGTGATGATACGCTCAGCTGATGAGTGAGAA
 ATATCGCGTTCGTGCCAGAGAGCGACGTTTTTCATAAGCCGTAATCATGTGACCACGAATG
 ACACGCGCCAGACCAGTCATATTTTCAGAACCGATTGGGTTGCGTTTGTGAGGCATTGCT
 GAAGACCCTTTTGCCTTTAGCAAAGAACTCTTCTACTTCGCGTTGCTCAGATTTTTGT
 AGACCACGAATCTCAGTCGCCATACGTTTCGATTGAAGTCGCAATGCTGGCAAGAACCGCA
 AAGTACTCAGCGTGAAGGTCACGAGGAAGGACTTGTGTTAAAGATTCCTTGGGCACGGAT
 GCCAAGATTTATCGCAGACATACTCCTCTACAAATGGTGGGATATTGGCAAAGTTCCCAA
 CCGCACCAGAAATCTTACCAGCTTCTACACCAGCAGCCGCATGCTCGAAGCGCTCGATAT
 TGCCTTTCATTTTCGCTGTACCAAGTTGCTAATTTAAGACCAAAGGTTGTCGGCTCAGCGT
 GCACACCATGAGTACGCCCCATCATGATGGTGAACCTTGTGCTCCTTGGCCTTGTGACGGA
 TGATATTAGTGAAGTTTTCAAGGTCACGACGGATGATGTCGTTGGCCTGCTTGTAGAGGT
 AACCATAAGCAGTATCCACCACGTCGGTAGAAGTTAACCCTAGTGAACCCACTTGCGCT
 CTTACCAAGAGTCTCAGAAACCGCACGCGTGAAAGCCACCACATCGTGGCGCGTCTCCT
 GCTCAATTTCCAAAATACGGTCGATGTCAAAGTCCGCCTTCTTGCGAATCAAAGCCACAT
 CTTCTTAGGGATTTCCCCAACTCAGCCCATGCCTCGTCAGAGAGGATTTCCACCTCAA
 GCCAAGCACGGTATTTATTTTCTTCACTCCAAATATTTCGCCATCTCAGGGCGAGAGTAAC
 GGTGATCATGTGTTAATTTTCTTCTTCTTAAGAT

ORF Predictions:

ORF #	Start	End	Direction	Length
1	930	1616	R	229 aa

>[SEQ ID NO:142] 3864180-2 ORF translation from 930-1616, direction R
 VPKESLTQVLPRDLHAIEYFAVLASIATSIERMATEIRGLQKSEQREVEEFFAKGQKGSSA
 MPHKNRPIGSENMTGLARVIRGHMITAYENVALWHERDISHSSAERIITPDTTILIDYML
 NRFGNIVKNLTVFPENMIRNMNSTFGLIFSQRAMLTLIEKGMTREQAYDLVQPKTAYSWD
 NQVDFKPLLEADSEVTSRLTQEEIDEIFNPVYYTKRVDDIFERLGLGD*

Description:

ADENYLOSUCCINATE LYASE (EC 4.3.2.2) (ADENYLOSUCCINASE) (ASL). -
 BACILLUS SUBTIL IS.

Assembly ID: 3864184
Assembly Length: 4392bp

>[SEQ ID NO:50] 3864184 Strep Assembly -- Assembly id#3864184
CCCTTTTGCCTCTCCCTTTGGTGCAGATTCTTTTGGGAATTGTGATTGGTCTCTTTTAC
CCAATACTGACTTTCATCTTAATACGGAGTTGTTTTTGGCCTGGTTATCGGACCCCTTGCT
TTTCCGAGAGGCTGAAGAAGCAGATGTTACGGCTATTTTAAACACTGGCGAATCATTGT
TTATCTCATATTTCCAGTGATTTTATCTCGACCCTGAGTTTGGGTGGCTTGGCCCATCT
TCTTTGGTTCAGCCTTCCCTTGGCAGCTTGCTTGGCTGTTGGGGCAGCCCTTGGTCCTAC
GGACTTGGTGGCCTTTGCCTCTCTTTCGGAGCGTTTATAGCTTTCCTAAGCGGTGTCCAA
TATTCTTAAGGGCGAAGGACTCTTGAATGATGCTTCTGGTTTGGTGGCTTTTCAGGTAGC
TTTGACAGCTTGGACAACCTGGAGCTTTTTCTCTGGGGCAAGCTAGCAGTTCGCTCATCTT
TTCAATCCTAGGCGGTTTTTTAATTGGATTTTTTAACAGCCATGACCAACCGCTTCCTCCA
TACCTTCTTGCTAAGTGTGCGCGCAACGGATATTGCCAGTGAACCTTTATTAGAATTGCA
GTTTGCCTCTAGTGACCTTCTTTCTGGCAGAAGAAGTCCATGTTTCAGGGATTATTGCCG
TCGTAGTTGATCGAATTTTAAAGGCAAGTCGCTTCAAGAAAATCACGCTCCTCGAAGCCC
AAGTGGATACGGTGACCGAGACGGTCTGGCATAACAGTGACCTTTATGCTCAACGGTTCTG
TCTTTGTGATTTTAGGGATGGAGTTGGAATGATAGCAGAACCTATCTTGACCAATCCAA
TCTATAATCCTCTACTTTTATTGCTATCTCTCATCGCCCTTACCTTTGTCCTCTTTGTCA
TTCGTTTTTATTATGATCTATGGCTATTATGCCTATAGAACCCGACGCCCTAAAGAAAAAGC
TAAATAAGTATATGAAGGACATGTTTCTCTTGACCTTTTCAGGTGTTAAGGGAACGGTGT
CGATTGCTACGATTCTCTTGATACCAAGTAATCTAGAACAGGAGTATCCTCTCTTGCTTT
TCCTTGTTGCAGGTGTGACGCTTGTGAGCTTTTAAACAGGTCTCTTGGTCTTGCCTCATC
TTTCTGATGAAGAGGAAGAAAGCAAGGATTATCTCATGCATATCGCCATTTTGAATGAAG
TAACGCTAGAGTTGGAAAAAGAGTTGGAAGACACCAGAAATAAACTTCCCCTCTATGCGG
CTATTGACAATTCGATCATGGACGTATTGAAAATCTCATTTTAAAGCCAAGAAAACCAGGA
TGATCAAGAAGACTGGGCTGCTTTGAAAATCGAATTCTTAGTATTGAAAGTGATGGTTTG
GAACAGGCCTATGAAGAGGGGAACATTAGCAATCGTGCTTACCGAGTTTACCAACGTTAT
CTGAAAAATATAGAACAAGGAATCAATCGTAAACTTGCCTCAAGACTGACCTATTATTTT
CTTGTTTCCTTGAGGATTTTACGTTTCTTCTTCATGAAGTTTTTACTCTTGGAAGACC
TTCCGTAGCTGGAAGGACAAGGAGCAAAGCCGTCTCCGTGCTCTTGATTATGACCAAATT
GCAGAGCTCTATCTTGCCAATACAGAGATGATTATTGAAAGTTTGGAACCTGAAGGGA
GTCTACAGACGCTCTTTGATTAGTTTTATGCAGGAGTCTCGTCTTCGAGAAACAGCTATT
ATCAGCAGTGGTGCCTTTGTGCAACGGGTATCAATCGTGTCAAACCAACAATATCGAT
GAAATGCTGAGAGGCTATTATCTGGAGCGCAAGTTGATTTTGAATACGAAGAAAACGA
TTGATTACGACTAAGTATGCCAAGAAATTACGACAAAATGTAAATAACTTAGAGAACTAT
TCCTTGAAGGAAGCTGCCAATACCCTGCCGTATGATATGGTGGAAATTGGTAAGAAGAAAT
TAGTTAATACTCTTCGAAAATCTCTTCAAACCACGTCAGCGTCGCCTTGGATTATATATG
TGACTGACTTCGTCAGTTTCATCTACAACCTCAAAGCAGGGCTTTGAGCAACCTGCGGCT
AGCTTCCTAGTTTGCTCTTTGATTTTCATTGAGTATAAGATTGTAAGTGAAGGAGTGTA
CATGAAAAAATGGGGAAAGAGCCTGAACTAGTCTGTCTACTTTTACCCAATCACACTTC
CATTTGGTACAGCTGGATCAACTGTGAGAAGGGATCGAATTTGCCATCATGTTACAGCTGA

GAGAATCATACCCTGGCTGACATATTTTTTCATCATTTTACGTGGTTTGAGGTTAGCAAC
 GATTTGAACTTTCTTGCCGACCAATTCTTGTTTCATTTGGATAGTATTTTGCAATTCCTGA
 AAGAATCTGACGATCTTCTCCATCACCAGCATCCAAGCGGAATTGAAGCAACTTATCTGA
 ACCTTCTACTTTAGACACTTCTTTGACTTCTGCGACACGGATTTCAACCTTGTCAAAGTC
 TTCAAACCTTGATTTTCATCCTTGTTTAGTTTGAGCTCAACTTCGTCCGGATTCCATTCTTT
 TTCGACTGCTGGTTTATTGCCTTCCATTTGTTTCCTTGATATAGGCGATTTCTTCTTCCAT
 ATTTAGACGTGGAAAGATAGGTGTTCCCTTTGGCAACTACAGTCACATCTGCTGGGAAGTC
 AGCCAAACTCAAGTTTTCAAGACTAGAACTTCTTCCAAACCAAGTTGAGTCAAAACTGC
 ACGACTAGTTTCCATCATAAATGGTTCAATCAAGTGAGCAACTACACGAATGCTGGCTGC
 CAAGTGGCTCATGACACTTGCCAATTGGTCACGAAGAGCTTCATCCTTGTCGAAGACCCA
 TGGTGACGTCTCATCGATGTATTTATTGGTACGAGAGATCAGAGTCCAGACTGCTTCAAG
 CGCACGTGGATAGTCAACTGCTTCCATGTGTGTATGGAAGTCTGCGATTGATTTTTCTGC
 AACCTCAGCAAGAACATGATCAAATTCAGTCACACCTTCTACATAGGCAGGGATTTGTCC
 ATCAAAGTACTTATTAATCATGGAAACCGTACGGTTAAGGAGGTTCCCAAGGTCATTAGC
 CAATTCATAGTTGATACGACCGACATAGTCTTCAGGAGTAAAGGTTCCGTCTGAACCAAC
 TGGAAGGTTACGCATGAGGTAGTAACGAAGTGGATCTAGTCCATAACGCTCTACCAACAT
 TTCAGGGTAAACGACATTCCCTTTTGACTTAGACATTTTTCCGTCTTTCATGACAAACCA
 ACCATGGGCAATCAAACGATCAGGTAATTTAACATCCAACATCATAAGAAGGATTGGCCA
 GTAGATAGAGTGGAAGCGAAGGATGTCTTTTCCCTACCATATGGAAGACTGTTCCATTCCA
 GAACTTGTCAAAGTTACCATGTTTCGTCTTGAGCGTAGCCAAAAGCTGTGCGATAGTTAAG
 AAGGGCATCAATCCAAACGTAGACAACGTGTTTTGGATTGTGATGGGACAGGCACTCCCCA
 TGTAAGGTTGTACGAGATACCGCCAAATCTTCCAAACCTGGCTCGATGAAGTTGCGTAG
 CATTTTCATTAAGACGACCATCTGGCGTGATAAATTCAGGATGAGCTTTGAAAAATTGAC
 CAAACGGTCTTGGTATTTGCTAAGGCGAAGGAAGTATGATTCTTCAGAAACCCATTCAAC
 CTCATGACCTGATGGAGCAATACCACCAGTCACATTTCCAGCTTCATCACGGAAAACCTTC
 TGCCAGCTGGCTTTCTGTAAAGAATTCTTCGTCTGATACTGAATACCAACCAGAGTATTC
 ACCCAAGTAGATATCATCTTGAGCAAGTAAGCGTTCAAAGACCTGTGCGACAACCTTTTTC
 ATGGTAGTCATCGGTTGTACGGATAAATTTATCGTATGAGATATCTAGTAATTGCCAGAG
 TTCTTTAACTCCAACCGCCATTCCATCAACATAGGCTTGAGGTGTAATACCAGATTCGAA
 TTCCGCTTTCTGCTGGATTTTCTGACCATGTTTCATCAAGACCTGTCAGATAAAATACATC
 GTAGCCCATCAGGCGTTTGTAACGTGCTAGGACATCACATGCGATAGTTGTGTAGGCAGA
 ACCGATATGAAGTTTCCCAGATGGATAGTAAATCGGCGTTGTAATATAAAAATTTTTTTC
 AGACATAATTTTTCTTTCCAGGCAAATGAAACCTGTTTTTCTAACACTTCATTATATCA
 CATTTTTAATGAATTTTCGATAGGGAAATCCATACCAAACAAGATAGACGAGTGTCCATC
 TTGTTGATCTCATTCATAACGAAGGGCTTCAATTGGATCAAGTTTCGATGCCTTGTTGGC
 TGGCAAGACTCC

ORF Predictions:

ORF #	Start	End	Direction	Length
1	197	670	F	158 aa
2	612	1304	F	231 aa

>[SEQ ID NO:143] 3864184-1 ORF translation from 197-670, direction F
VIFISTLSLGGLAHLLWFSPLAACLA VGAAALGPTDLVAFASLSERFSF PKRVSNILKGE
GLLNDA SGLVAFQVALTAWTTGAFSLGQASSSLIFSILGGFLIGFLTAMTNRFLHTFLLS
VRATDIASELLLEFEFASSDLLSGRRSPCFRDYCRRS*

Description:

unknown

>[SEQ ID NO:144] 3864184-2 ORF translation from 612-1304, direction F
VTFFLAEEVHVSGIIAVVVDRI LKASRFKKITLLEAQVDTV TETVWHTVTFMLNGSVFVI
LGMELEMIAEPILTNPIYNPLLLLLSLIALTFVL FVIRFIMIYGYAYRTRRLKKKL NKY
MKDMFLLTFSGVKGT VSIATILLIPSNLEQEYPLLLFLVAGVTLVSFLTGLLVLP HLSDE
EEESKDYL MHIAILNEVTLELEKELEDTRNKLPLYAAIDNSIMDV LKISF*

Description:

unknown

Assembly ID: 3864194

Assembly Length: 1941bp

>[SEQ ID NO:51] 3864194 Strep Assembly -- Assembly id#3864194
AATTAGTATTCTCAACCTTTTTTATCTTGATAGTTCAAGATGGCATTCGTTGAATTGGTAA
CATAGTA ACTATCCACTCCCTTCAGTTTAGCTGCCTCTTGAACCCAGGATTCTTGCGGTT
TTGGCGGTTCAACAGGAATTCTTTTTCTTTTCCAGAAACCGTAAAAGCTGATTGTTTCTG
AGTAAAAGACCCATCTTTACTTTTTTTAGGAGAGAAAAAGACGCTAATATTTTTCTGAGA
TTTAGTCATATCTTTATTGACTTGACGAGATAGGGAATCACCCAAAGCCATAATCACAAAC
AACTGATGAAACACCGATAATAATCCCAATCATAGTAAGCAAAGAACGCATCTTG TGAGC
CATGATAGATGAAAAGGCAAATTT CAGATTCTGCATCTTAGTTTTCTCCTTTCCTAACT
GAGCACTGTCAGACGAAATGACCCCATCCCGAATGACAATCTGACGTTTGGCATAGGCAG
CAATCTCAGGCTTCATGCGTTACCATGATAATGGTTTTCTTCTTTATTCAAATCAACC
AATAATTGCATAATTTGGTTACCTGTTTTGGTATCCAAGGCTCCTGTCGGTTCATCCGCT
AGGATAATAGAAGGATTGTTTACCAAGGCACGCGCAATGGCTACACGTTGCTTTTGACCA
CCAGATAATTCTGAAGGTAAATGGTGACTACGTTCTATCAATTCAACCTTGTCTAAATAT
TCCTCAGCCAACTTGCGACGTTTTGAAGACGAAACTCCTGCGTAAATCAAGGGCAATTCT
ACATTTTGCAGAGCATTGAGCTTCGATAGAAGAAAGAACTGCTGAAAGACAAAACCGATT
TGTTGGTTACGGACCTTAGCTAGTTGTTTTTACCAAGCCCAGCCACTTCTTGACCTTCA
AGATAATATTCTCCACTGGTTGGTGTATCCAACATGCCAATCGTATTCATCAGAGTGGAC
TTACCAGACCCAGATGGTCCCATGATGGCTACAAATTCACCCTCATTCACCTTCTAGATTG
ATATTTTTTGAGAACCTGCAGTTCTTGGTCACCATTACGGTAACTTCTGAAGATATTTTTT
AGACTAATTAGTTGCTTCATCAGCCTTCACCTCTTTTCTTCTTCCAAGGAAGATGTTGG
ATTACTGATGACCTTAGCACCGTTCGTTAAACCAGAAGTGATTTCTTGATTTTCTGCGTC
AGCATTTCCCAATGAAACCTCAACTTTTTTAGCCTTTTGTTGTTTCATCCACAATCCAGAC
ATAATTTTTACTATCATCCATTACTAGACTGCTAACAGGAACAAGAATAGCCTTAGTTTT

GCTTTTAACCTCAATGTTGACAGAAAAACCTTGTTTCAAATCACCAACCTCGCCTGTAC
 ATCAATAGTATAAGGGTATTTAGAACCTGTATTATTCCCGGCTGCTGGACTAGCTGCTTC
 ACCATTGTTTTTAGGATAGTCAGAAATATAGGCTTAATTTCCAGTCCATTTTTTATCAG
 GATACTTTAGAAAGTAAAGCTTACTTCTTGACCTACAGAAAGGTTGGCTAGATTGTACT
 CAGACAATTCTCCCTTGACTTGTAATTTTCATTGCTGACAATATGAACCATAACTTGAC
 TCGCCCCGTGTGGAGATTTAGAAACATTGCTATTGACTTCGACTACAGTTCCCTCTAGGG
 TACTGAGAACAGTTGTTGCATCCAATTGACTTTGAGCCTTGCTTAATTGCGCTGCAGCAT
 CTGCACGCGCATCACGGGCATCACCCAATTGAGCATCAATAGAAGCAACAGAAATTTCCAG
 CCACTGGAGTTGGGCTTTGCACCGTTGCATCTTCTCCTCCTACTGGCGCTGGTAAGTGTG
 GAGCCTGAGCTGAAGCGGCTTCATTTCTGCTTGATTGAGTTCATTGATATGACGATCTG
 CCTTAGCTACTGCTCGACTAG

ORF Predictions:

ORF #	Start	End	Direction	Length
1	1084	1380	R	99 aa

>[SEQ ID NO:145] 3864194-3 ORF translation from 1084-1380, direction R
 VTGEVGLDKQGSVNIEVKSSTKAILVPVSSLVMDSDKNYVWIVDEQQKAKKVEVSLGNA
 DAENQEITSGLTNGAKVISNPTSSLEEGKEVKADEATN*

Description:

unknown

Assembly ID: 3864338

Assembly Length: 1335bp

>[SEQ ID NO:52] 3864338 Strep Assembly -- Assembly id#3864338
 ATCGAATTCCTATTTTAACACTTTCTTTTCTAAAACAGTCTATATTTTATTTCAAACTG
 TATTATATTTTGAATAAAGTCTTTTCTTTTTCAGAAAAAGGGTATAATA
 AAAGAAAAAAGCAGTAACACTCAATGGAATCGAAAAAGCAAAGCTAGGAAGCTAGCCGC
 AGATTGCTCAAAACACTGTTTTGAGGTTGCAGATAGAGCTGACGTGGTTTGAAGAGATT
 TCGAAGAGTATAAAAAGGTGCTAGGCATGTTGATTTTCTTTTGTAAATGATTGTCAA
 GAAAAATCATCCATATTGGACATGGATGCCTTTTTTGCTGCAGTGGAAATCAGGGATAAT
 CCTAACTCAGAGGAAAACCTGTCATTATTGGAAGCGACCTCGGCAAACAGGTGGACGG
 GGAGTCGTTTCTACCTGTAGTTATGAGGCAAGAGCTTTTGGTGTCCATTCTGCCATGAGT
 TCCAAGGAAGCTTATGAACGTTGTCCCAGGCTGTCTTTATCTCAGGGAATTCGATGAGA
 AATACAAGTCTGTGGGACTCCAGATTCGAGCTATTTTAAAGCGCTATACAGATTGATTG
 AACCCATGAGCATTGACGAAGCCTATTTGGATGTGACAGAAATAAACTCGGTATCAAGT
 CAGCGGTCAAATTTGCTCGCCTCATTCAAAAAGATATCTGGCAAGAACTCCATCTAACTG
 CTTCGCGAGGCGTTTCTTACAACAAATTTAGCTAAAAATGGCGAGTGATTATCAAAAAC
 CACATGGTTTGACAGTGATTCTACCTGAACAGGCTGAGGATTTTCTCAAACAAATGGATA
 TTTCCAAATTTTCATGGAGTAGGAAAAAGACAGTAGAACGTCTTCATCAAATGGGCGTTT

TTACTGGTGCTGATTTACTTGAAGTTCCTGAGGTAACCCCTAATAGACCGTTTTGGTAGAC
 TAGGCTATGATCTGTATCGAAAGGCTCGTGGCATTTCACAACTCTCCAGTCAAATCCAATC
 ACATCCGTAAATCAATCGGCAAGGAGAAAACCTACGGGAAGATTCTCCGTGCTGAGGAAG
 ATATCAAAAAAGAGAGCTGACTCTTCTATCAGAAAAAGTCGCTCTCAATCTACATCAACA
 AGAAAAAGCTGGAAAAATTGTCATTTTGAAAAATCCGCTACGAGGACTTTTCAACTCTTAC
 CAAACGAAAAAGTATTGCTCAAAAAACACAAGATGCTAGTCAGATAAGCCAAATAGCCCT
 GCAACTCTATGAAGAATTAAGTGAGAAAGAAAGAGGTGTCCGCCTATTGGGGATTACCAT
 GACTGGATTTTTAAAG

ORF Predictions:

ORF #	Start	End	Direction	Length
1	552	1100	F	183 aa

>[SEQ ID NO:146] 3864338-2 ORF translation from 552-1100, direction F
 VGLQIRAIKRYTDLIEPMSIDEAYLDVTENKLGKISAVKILRIQKDIWQELHLTASAG
 VSYNKFALAKMASDYQKPHGLTVILPEQAEDFLKQMDISKFHGVGKKTVERLHQMGVFTGA
 DLLEVPEVTLIDRFGRGLGYDLYRKARGIHNSPVKSNHIRKSIGKEKTYGKILRAEEDIKK
 ES*

Description:

ECODINJ NCBI - Escherichia coli (sub_strain W3110, strain K-12)
 DinP, DNA damage inducible protein

Assembly ID: 3864360

Assembly Length: 1796bp

>[SEQ ID NO:53] 3864360 Strep Assembly -- Assembly id#3864360
 TCCAAGCTAGCTATTTCTGTGGAAGGGGCTTCGGTTGGCAGAACCTGGTGAATTTACCCAA
 ACGTGCTTTTTTAAACGGTCGCGTAGACTTGACACAGGCAGAGGCTGTGATGGATATCAT
 CCGTGCCAAGACTGACAAGGCCATGAACATTGCGGTCAAACAATTAGACGGCTCCCTTTC
 TGACCTCATTAACAATAACCGTCAAGAAATCCTCAATACACTTGCCCAAGTTGAGGTCAA
 TATCGACTATCCTGAATATGATGATGTTGAGGAAGCTACTACTGCCGTTGTCCGTGAGAA
 GACTATGGAGTTTGAGCAATTGCTAACCAAGCTCCTTAGGACAGCACGTCGTGGTAAAAT
 CCTTCGTGAAGGAATTTCAACGGCTATCATTGGACGTCCCAACGTTGGGAAATCAAGCCT
 TCTCAACAACCTCTTGCGTGAGGACAAGGCTATCGTAACCGATATCGCTGGGACAACACG
 AGATGTCATCGAAGAGTACGTCAACATCAATGGTGTTCCTCTAAAATTGATTGACACAGC
 TGGTATTCGTGAAACGGATGATATCGTTGAACAAATCGGTGTTGAGCGTTGAAAAAAGC
 CCTCAAGGAAGCCGACTTGGTTCTACTAGTGCTAAATGCCAGTGAACCACTGACTGCGCA
 AGACAGACAACTTCTTGAAATTAGCCAAGATACCAATCGCATTATTCTACTTAATAAAAC
 CGACCTGCCAGAAACGATTGAAACTTCGAAACTACCTGAAGACGTTATCCGTATTTTCAGT
 CCTTAAAAACCAAAACATCGACAAGATTGAAGAGCGAATCAACAACCTCTTCTTTGAAAA
 TGCTGGCTTGGTCGAGCAAGATGCTACTTACTTGTCAAACGCCCGTCACATTTCCCTGAT

TGAAAAAGCAGTTGAAAGCCTACAAGCCGTTAATCAAGGTCTTGAGCTGGGGATGCCAGT
 TGATTTGCTTCAAGTTGACTTGACTCGTACTTGGGAAATCCTCGGAGAAATCACTGGGGA
 TGCTGCTCCAGATGAATCATCACCAACTCTTTAGCCAATTCTGTTTAGGAAAATAAGA
 AAAATCCATGATCCTTCATTTCGGTCATGGATTTTATTGTCTTTATTAGTAATCTGGTCTT
 AAGACCCCTGTTACAGTTGCCTTAGTTGCTTCGTAGTCGCCATCTACGACAACCTTGATA
 ATGCGTTTGACATCTTCTTCTGGTGCTGGAACAAGAGGTAGACGAGTGGGTCCAGCTTCA
 AATCCCATATAGTTAAGAATTGCCTTAAGTGGAGCAGGACTTGGATAAGAGAAGAGAGCA
 TTAACCTTAGGAATGAATTTACGCTGAATTGCTGCGGCTTTCTTCATATCGCTTTCTGCA
 ATGGCAGTAAACATCTCGTGCATTTTCATCCCCATTTGTATGAGAGGCAACAGAAATAACC
 CCATCCGCCCCCAAGGTTTCATGGCATGGAAAGCATCTCCATCCTCACCTGTATAAATCAAG
 AACTCTTCAGGCTTGTGCTCAATCAAGTAAGCCATATTAGCCAAGCTAGTACATTCTTTG
 ACACCGATAATATTTGGATGGTCAGCCAAGCGAAGCATGGTTTCTGGAGTCAATTTCGACA
 ACTACACGCCCTGGAATGTTATAGATAATAATTGGTAGGTCAGAAGCATCTGCAATAGCC
 TTAAGTGCTGATACATCCCTTCTTGAGAAGGTTTGTGTAGTAAGGAACAATAGCAAGC
 CCAGCTGCGAAACCACCAAATTCGCTACTTCTTTGACAAACTCAATAGAGTCACG

ORF Predictions:

ORF #	Start	End	Direction	Length
1	47	1078	F	344 aa

>[SEQ ID NO:147] 3864360-1 ORF translation from 47-1078, direction F
 VNLPKRAFLNGRVDLTQAEAVMDIIRAKTDKAMNIAVKQLDGSLSDLINNTRQEILNTLA
 QVEVNIDYPEYDDVEEATTAVVREKTMEFEQLLTKLLRTARRGKILREGISTAIIGRPNV
 GKSSLLNNLLREDKAIVTDIAGTTTRDVIEEYVNINGVPLKLIDTAGIRETDDIVEQIGVE
 RSKKALKEADLVLLVLNASEPLTAQDRQLLEISQDTNRIILLNKTDLPETIETSKLPEDV
 IRISVLKNQNIIDKIEERINNLFENAGLVEQDATYLSNARHISLIEKAVESLQAVNQGLE
 LGMPVDLLQVDLTRTWEILGEITGDAAPDELITQLFSQFCLGK*

Description:

THIOPHENE AND FURAN OXIDATION PROTEIN THDF. - ESCHERICHIA COLI.

Assembly ID: 3864388

Assembly Length: 2337bp

>[SEQ ID NO:54] 3864388 Strep Assembly -- Assembly id#3864388
 CTTTCGTACAGGTGGTTCCTATGCAAGGGTGGAAGCCAATCGTCAGAACAACAAGCATCTT
 CATCAAGCCAGAACTGGAGCAATTACAAAAAGAAATTGCTGAAGAAGAAGCAAGCTTGGG
 TTCAGAAGAAGTGGCTTTGAAGACCTTGCAAGATGAGATGGCCAGATTGACCGAGTCATT
 AGAAGCTATTAAATCTCAAGGAGAGCAGGCACGTATTCAGGAGCAAGGCTTGTCCCTCGC
 TTATCAGCAAAGTAGTCAGCAAGTTGAAGAACTGGAAACTCTTTGGAAACTCCAAGAAGA
 GGAAATAGATCGTCTTTCCGAGGGAGATTGGCAAGCGGATAAGGAAAAATGCCAAGAGCG
 TCTTGCTGCAATCGCCAGTGACAAGCAAAATCTGGAAGCTGAGATTGAAGAGATTAAAGTC

TAATAAAAATGCCATCCAAGAACGCTATCAAAACTTGCAGGAAGAGCTAGCGCAAGCTCG
 TTTGCTTAAGACAGAACTGCAAGGGCAAAAACGTTATGAAATTGCTGATATTGAACGCTT
 AGGCAAGGAATTGGACAATCTTGATTTTGAACAAGAGGAAATCCAGCGCCTTCTTCAAGA
 AAAGGTTGACAATCTTGAGAAGGTTGATACAGAAATTGCTCAGTCAACAGGCGGAAGAATC
 CAAAACCTCAGAAAACGAACCTCCAACAAGGTTTGATTTCGCAAACAGTTTGAGTTGGATGA
 TATAGAAGGTCAGCTGGATGATATTGCTAGTCATTTGGATCAGGCTCGCCAGCAGAATGA
 GGAGTGGATTTCGCAAGCAAACACGTGCTGAAGCTAAGAAAGAAAAGGTCAGCGAGCGCTT
 TGCCGCCATCTACAAAGTCAATTAACAGACCAGTACCAGATTAGCCATACTGAAGCTCTA
 GAAAAAGCGCATGAATTGGAAAACCTCAATCTGGCAGAGCAAGAAGTTAAGGATTTAGAG
 AAGGCTATTTCGCTCACTGGGTCCTGTCAATATAGAAGCTATTGACCGGTACGAAGAAGTT
 CACAACCGTCTGGACTTTTCTAAATAGTCAGCGAGATGATATTTTGTTCAGCGAAAAATCTG
 CTCCTTGAAACCATTACAAAGATGAATGATGAGGTTAAGGAACGCTTTAAATCAACCTTT
 GAAGCTATTTCGTGAGTCCTTTAAAGTGACCTTCAAGCAGATGTTTGGCGGAGGTCAGGCA
 GACTTGATATTGACTGAGGGCGACCTTTTACAGCTGGTGTGGAGATTTCTGTTCAACCTC
 CAGGTAAGAAAATCCAGTCGCTTAACCTCATGAGTGGTGGTGAAAAAGCCCTATCGGCTC
 TTGCCTTGCTTTTCTCCATTATTTCGTGTCAAGACCATTCTTTTGTTCATCTTGGATGAGG
 TGGAAGCTGCGTTGGATGAAGCCAATGTTAAACGTTTTTGGGGATTACCTCAACCGCTTTG
 ACAAGGACAGCCAGTTTATCGTCGTAACCCACCGTAAGGGAACCATGGCAGCGGCCGATT
 CCATCTATGGAGTGACCATGCAAGAATCGGGTGTTCAAAGATTGTTTCAGTTAAGTTAA
 AAGATTTAGAAAGTATTGAAGGATGACAATTAACTAGTAGCAACGGATATGGACGGAAC
 CTTCTTAGATGAGAATGGGCGCTTTGATATGGACCGCCTCAAGTCTCTCTTGGTTTCTTA
 CAAGGAAAAAGGATTTACTTTGCGGTGGCTTCGGGTGCGGGATTTCTGTCTCTGGAAAT
 CGAATTATTTGCTGGTGTTCGTGATGACATTATTTTCATCGCGGAAAATGGCAGTTTGGT
 AGAGTATCAAGGTCAGGACTTGATGAAGCGACTATGTCTCGTGACTTTTATCTGGCAAC
 TTTTGAAAAGCTGAAAACGTCACCTTATATAGATATCAATAAACTGCTCTTGACGGGTAA
 GAAGGGTTCATATGTTCTAGATACGGTTGATGAGACCTATTTGAAAGTGAGTCAGCATTAA
 TAATGAAAATATCCAAAAAGTAGCGAGTTTGGAAAGATATCACAGATGACATTTTCAAATT
 TACAACCAACTTCACAGAAGAAACGCTAGAAGCTGGTGAAGCTTGGGTCAATGATAATGT
 CCCTGGTGTCAAGGCTATGACAACTGGCTTTGAATCTATTGATATTGTTCTGGACTATGT
 CGATAAGGGTGTAGCTATTGTTGAATTAGCTAAAAAACTTGGCATCACAATGGATCAGGT
 CATGGCTTTTGGAGACAATCTTAATGACTTACATATGATGCAGGTTGTGGGACATCCTGT
 AGCTCCTGAAAATGCACGACCAGAGATTTTGAATTAGCATAAGACTGTGATTGGTC

ORF Predictions:

ORF #	Start	End	Direction	Length
1	1239	1586	F	116 aa

>[SEQ ID NO:148] 3864388-3 ORF translation from 1239-1586, direction F
 VEISVQPPGKKIQSLNMSGGEKALSALALLFSIIRVKTI PFVILDEVEAALDEANVKRF
 GDYLNRFDKDSQFIVVTHRKGTMMAADSIYGVTMQESGVSKIVSVKLKDLESIEG*

Description:

P115 protein - *Mycoplasma hyorhinitis* (SGC3) (similarity to SMC1_YEAST, chromosome segregation protein)

Assembly ID: 3864406

Assembly Length: 2162bp

```
>[SEQ ID NO:55] 3864406 Strep Assembly -- Assembly id#3864406
CTAAAAGTGAAGCCCGATAGCGTCTCTCTCCTGCAAGGATTCATAACCAATAACAGGAG
ATTGACGAACAATAATCGGTTGAATGACCCCATTTTCTTTGATAGACTGTGCTAGTTCAT
CTAGCTTTTCTCTATCAAATTCTTTTCGGGGTTGATAGGGATTTTTTTGTATATCTGTGA
TAGAAATCATTTCAAATTTTTCCATGATTCTACACTAACACATCTTTTCTCTTATGTAAA
GCTTTCTTTACATAGATGTCAATTAAGATTCTAAATCACCTGAACTCTTGTTAAGTTTGA
TAGAGGTAGTTTCTTCTTTCCCGTTACGATAGTAGGTTATCTTAATGGTGTCTCCGATAG
AATGGTTGTAAAGAGCACTTTGTAAGTCTGTTGATGAAGCAATCTCTTTGTCTACTT
TTGTAATTACATCGTATTTTTCAAGGTGACCATTGGCAGGCATATTACTTTGTACCGAAC
GAACAATTACACCAGATGTAACATTACTTGGAATATTGAGTCTTCTGATGTCGCTTGTA
TCACATTAGATAAATTAACCATCTGGATTCCCAAAGCTGGACGCGTCACTTTTCCGTTTT
TTTCTAACTGTTCAATAATATTGATAGCATCATTTGCAGGAATTGCGAAACCAAGACCTT
CTACAGATGTTCTCCATTTGTAGCAATTTTACTTGAGGTAATTCCGATAACCTGCCCTT
GAATATTGATCAGTGGGCCGCCAGAGTTACCTGGGTTAATAGCAGTATCAGTTTGGATGG
CTTTTGTAGAAATAGCTTGTCATCTTCCGATTTTAAGGATACATTTCTATTGAGACTGG
ATACGATACCTTGAGTGACAGTATTTGCATATTCAGAACCTAACGGGCTACCGATGGCAA
TAGCAGTTTCTCCTACAGTTAACTTACTAGAATCACCAAACCTCAGCTACTGTTGTCACTT
TTTCTGAAGAGATTTTCGACGACAGCAATATCAGAGAAAGTGTGAGCTCCGACAATTTCTC
CAGGTACTTTAGTCCCATCTGACAATCGAATATCTACTTTGCTGGCGCCATTTATAACGT
GATTGTTGGTGACGATGTAAGCTTCTTTATCATTCTTTTTATAAAATAACTCCAGATCCTT
CACTAGAGATTTCGCTGAGAATCTGTGTGAGTATCATCATTGCCAAATACGCTATTTTGTG
TGTTTGCCGAATAAGTAATAACAGAAACAACAGCATCTTTTACTTTGTTAACGGCCTGTG
TTGTTGAATTTTCCGTTCCCTTATAGGCAGTTTGTGTAATAGTACTATTGTTGTTAGAGTT
GTTTACACTACTTTTTTGGAGTTAGTTGAGTTATTGAAAACTACCCAAGGCTCCACTAAA
AAAGCTAATGACGATAACGACTAATAATTGAAACCATTTTTTGTAAAATGTTTTTAGATG
TTTCATATTTGCCTCCATATGTTTGAATTACTGAAAGTATAAACTGACTAGCTTAATTAT
AACTTAAACACAAAAGTTTTACACAACTGTGGATAACTCTTTTGAAGTGTGATTTTCT
TAATTGAAATCTATTTTTTATTTTGTGAATAAGATGTGAAAAATAGAGAATATGTTAGA
ATAGAGTCATGAAAAATTAAAGTTGTAACAGTTGGGAACTGAAAGAAAAGTATTTAAAG
ATGGTATCGCAGAGTATTCAAACGAATTTCTAGATTTGCTAAGTTTGAAATGATTGAGT
TATCAGATGAAAAAACACCAGATAAGGCCAGTGAATCAGAAAATCAAAGATTTTAGAAA
TAGAAGGTCAGAGAATTTTATCAAAAATTGCTGACCGTGATTTTCGTTATTGTGTTAGCCA
TTGAAGGGAAAACCTTCTTCTCAGAAGAAATTTAGTAAGCAGTGAGAAGAACTTCTATAA
GGAAGGATGTCTACTCTTACTTTTATTATTGGGGGAAGTTTAGGATTGTCATCATCTGTA
AAAAATAGAGCCAATCTTCTGTGAGTTTGGTTCGCTAACCTTGCCTCATCAGTTAATG
AGACTAGTTCTTGTGTAACAAATCTATCGCGCTTTTACGATTCAGCAGGGATTCCCCTAC
CATAAATAGAGAATTGACTTTTAATTGAATTTTTGGTAGAATAATTGTGTTAGGTCTCAT
```


AG

ORF Predictions:

ORF #	Start	End	Direction	Length
1	263	958	R	232 aa

>[SEQ ID NO:149] 3864406-1 ORF translation from 263-958, direction R
 VTTVAEFGDSSKLTVGETAIAIGSPLGSEYANTVTQGIVSSLNRNVSLKSEDGQAISTKA
 IQTDTAINPGNSGGPLINIQQGVIGITSSKIATNGGTSVEGLGFAIPANDAINIIEQLEK
 NGKVTRPALGIQMVNLSNVSTSDIRRLNIPSNVTSGVIVRSVQSNMPANGHLEKYDVITK
 VDDKEIASSTDLSALYNHSIGDTIKITYYRNGKEETTSIKLNKSSGDLES*

Description:

Bacillus subtilis (strain 168,) DNA. Homologous to E. coli serine
 protease HtrA (BLAST)

Assembly ID: 3864452

Assembly Length: 1766bp

>[SEQ ID NO:56] 3864452 Strep Assembly -- Assembly id#3864452
 ATCGAATTTTCCAAAATGGGGAGCTAGAGCAGTGGAGTGATTATGTGGCAGACGATTTGA
 TTCAGCATAATCATGAGATTGGACAAGGAAGTGCTGCTTATAAAACTATGTGGCTGAAT
 ATATTGTCACCTTTGACTTCGTTTTCCAACCTCTAGGACAAGGAACTATGTGGTTAGCT
 ATGGTCAGACTCAGATTGATGGCGTTGCTTATGCCAAGTACGATATCTTCCGTTTAAAGA
 ACGGGAAAATTGTGGAGCATTGGGATAATAAGGAAGTCATGCCTAAGGTAGAAGACTTGA
 CCAATCGAGGGAAGTTTAAATTGAGGACAAAGAATGATTGAATACAAAATGTAGCACT
 GCGCTACACAGAAAAGGATGTCCTTGAGAGATGTCAACTTACAGATTGAGGATGGGGAATT
 TATGGTTTTAGTAGGGCCTTCTGGGTCAGGTAAGACGACCATGCTCAAGATGATTAACCG
 TCTTTTGGAAACCAACTGATGGAAATATTTATATGGATGGGAAGCGCATCAAAGACTATGA
 TGAGCGTGAACCTCGTCTTTCTACTGGTTATGTTTTACAGGCTATTGCTCTTTTCCAAA
 TCTAACAGTTGCGGAAAATATTGCTCTCATTCCTGAAATGAAGGGGTGGAGCAAGGAAGA
 AATTACGAAGAAAACAGAAGAGCTTTTGGCTAAGGTTGGTTTACCAGTAGCCGAGTATGG
 GCATCGCTTACCTAGTGAATTATCTGGTGGAGAACAGCAACGGGTCCGGTATTGTCCGAGC
 TATGATTGGTCAGCCCAAGATTTTCCTCATGGATGAACCCTTTTCGGCCTTGGATGCTAT
 TTCGAGAAAACAGTTGCAGGTTCTGACAAAAGAATTGCATAAAGAGTTTGGGATGACAAC
 GATTTTTGTAACCCATGATACGGATGAAGCCTTGAAGTTGGCGGACCGTATTGCTGTCTT
 GCAGGATGGAGAAAATTCGCCAGGTAGCGAATCCCGAGACAATTTTAAAAGTGCCTGCAAC
 AGACTTTGTAGCAGACTTGTGTTGGAGGTAGTGTTTCATGACTAATTTAATTGCAACTTTTC
 AGGATCGTTTTAGTGATTGGTTGACAGCTACAATGACATTGGTCCGTTCCCTGAGCAAGA
 GATAGATTAGCCAGACAGTCATGCCCAAATCCCTCCAGGTAAGAGCATAGACCGTTGCA
 CATTAAGTACGATTAAAAAAGTGATAATGGCAAGAAACTTGCTACTGCTTGTAATAAAA
 AGGTTGTTAGTGTCATATTAGTTTCATCAATACCAAGGCGACAGAAGTTCCTGCCCTTAA

GCGAGGGTAATGAGCAGGGATTCAAACATCTTACTCATACCAGAGTTTATGTGGTTGGTC
 ATAATATCACGGACCGCATTGGTCAAGGCAATACCTGGTACAAACGGCATGACCGCACCA
 GCTATAATCAAATCTGCCGTTGAAGGAAAACCTGTGTAGCGAGCCCAAACTGGGCAATT
 ATCCCAAAGACAAAAGCTCCAGCAAAGGCTGTCAAAAGGGAATTCGGATAAAATTTTCC
 ACATAGAGGGAAAAGGCAAACCAAATAAGGTCGCCACTCCTGCCCCAAGTGCCTCGTAG
 ATATTTCCGCTAAACATAACTGAAAAGAAAGGAGCACTAAAGGTCGCAGCCAGAGTTACC
 TGCAACTTAGTATAGGGAAGGGGTTGAGCTTGCAAGGCCGTCAATTGCTTAAAGGCTGTT
 TCTAAGTCAATCTGCCCCCAACTGG

ORF Predictions:

ORF #	Start	End	Direction	Length
1	1079	1201	R	41 aa

>[SEQ ID NO:150] 3864452-2 ORF translation from 1079-1201, direction R
 VQRSMLLPGGILGMTVWLIYLLLEPTNVIVAVNQSLKRS*

Description:

unknown

Assembly ID: 3864458

Assembly Length: 1705bp

>[SEQ ID NO:57] 3864458 Strep Assembly -- Assembly id#3864458
 CTCTGACGGAGGCTGGTTATGTGGGTGAGGATGTGGAAAATATACTCCTCAAACCTTTGC
 AGGTTGCTGACTTTAACATCGAACGTGCAGAGCGTGGCATTATCTATGTGGATGAAATTG
 ACAAGATTGCCAAGAAGAGTGAGAATGTGTCTATCACACGTGATGTTTCTGGTGAAGGGG
 TGCAACAAGCCCTTCTCAAGATTATTGAGGGAAGTGTGCTAGCGTACCGCCTCAAGGTG
 GACGCAAACATCCACAACAAGAGATGATTCAAGTGGATACAAAAATATCCTCTTCATCG
 TGGGTGGTGCTTTTGATGGTATTGAAGAAATTGTCAAACAACGTCTGGGTGAAAAAGTCA
 TCGGATTTGGTCAAACAATAAGGCGATTGACGAAAACAGCTCATACATGCAAGAAATCA
 TCGCTGAAGACATTCAAAAATTTGGTATTATCCCTGAGTTGATTGGACGCTTGCCCTGTTT
 TTGCGGCTCTTGAGCAATTGACCGTTGATGACTTGGTTTCGCATCTTGAAAGAGCCAAGAA
 ATGCCTTGGTGAAACAATACCAAACCTTGCTTTCTTATGATGATGTTGAGTTGGAATTTG
 ACGACGAAGCCCTTCAAGAGATTGCTAATAAAGCAATCGAACGGAAGACAGGGGCGCGTG
 GACTTCGCTCCATCATCGAAGAAACCATGCTAGATGTTATGTTTGAGGTGCCGAGTCAGG
 AAAATGTGAAATTGGTTTCGCATCACTAAAGAACTGTTCGATGGAACGGATAAACCGATCC
 TAGAAACAGCCTAGAGGTGACTATGGAACCTAATACACACAATGCTGAAATCTTGCTCAG
 TGCAGCTAATAAGTCCCACTATCCGCAGGATGAACTGCCAGAGATTGCCCTAGCAGGGCG
 TTCAAATGTTGGTAAATCCAGCTTTATCAACACTATGTTGAACCGTAAGAATCTCGCTCG
 TACATCAGGAAAACCTGGTAAAACCCAGCTCCTGAACTTTTTTAACATTGATGACAAGAT
 GCGCTTTGTGGATGTGCCTGGTTATGGCTATGCTCGTGTCTTAAAAAGGAACGTGAAAA
 GTGGGGGTGCATGATTGAGGAGTAATTTAACGACTCGGGAAAATCTCCGTGCGGTTGTCA

GTCTAGTTGACCTTCGTCATGACCCGTCAGCAGATGATGTGCAGATGTACGAATTTCTCA
 AGTATTATGAGATTCCAGTCATCATTGTGGCGACCAAGGCGGACAAGATTCCCTCGTGGTA
 AATGGAACAAGCATGAATCAGCAATCAAAAAGAAATTAACTTTGACCCAAGTGACGATT
 TCATCCTCTTTTCATCTGTCTAGCAAGGCAGGGATGGATGAGGCTTGGGATGCAATCTTAG
 AAAAATTGTGAGGAAAAGAAAATGGCAAAAACAATTCATACAGATAAGGCCCCAAAGGCT
 ATCGGGCCCTATGTTCAAGGAAAAATCGTTGGCAACCTTTTGTGTTGCTAGCGGTCAAGTT
 CCCCTATCCCCTGAAACTGGGGAAATTGTAGGAGAGAATATCCAAGAACAGACAGAGCAA
 GTCTTGAAAAACATCGGTGCTATTTTGGCAGAAGCAGGAACAGACTTTGACCATGTTGTC
 AAAACAACCTTGTTTCTTGAGCGATATGAACGACTTTGTTCCTTTTAATGAGGTTTACCAA
 ACGGCCTTCAAAGAGGAATTCCCAG

ORF Predictions:

ORF #	Start	End	Direction	Length
1	797	1105	F	103 aa
2	1179	1391	F	71 aa

>[SEQ ID NO:151] 3864458-2 ORF translation from 797-1105, direction F
 VTME LNTHNAEILL SAANKSHYPQDELPEIALAGRSNVGKSSFINTMLNRKNLARTSGKP
 GKTQLLNFFNIDDKMRFVDVPGYGYARVSKKEREKWGCMIEE*

Description:

unknown

>[SEQ ID NO:152] 3864458-3 ORF translation from 1179-1391, direction F
 VQMYEFLKYYEIPVVIIVATKADKIPRGKWNKHESAIAKKLNFDPSSDFILFSSVSKAGMD
 EAWDAILEKL*

Description:

HYPOTHETICAL 22.0 KD PROTEIN IN LON-HEMA INTERGENIC REGION (ORFX). -
 BACILLUS S UBTILIS.

Assembly ID: 3864474

Assembly Length: 1673bp

>[SEQ ID NO:58] 3864474 Strep Assembly -- Assembly id#3864474
 ACGTTTTGGGAAGTTCGGATAGCAGATTCCGAACAACTGATAATGGTTGGCAAAATC
 ATTATTCCTAATAGTAACGAAGCTGGTTAGGACAACCTCATGCCATTCCTAAAAAGGTTT
 TAATCCAAGGCACCAATAATTGTAGGCCGAAAAAACCATAAACAATAGATGGAATGGCTG
 CCATCAAGTTGATAGCTGATTTTAAGAAGCTATAGACGGGCTTTGGACAATTATAAACCA
 TAAACACCGATGTCAAGATCGCCTGTTGGCACCCCAATCACAATCGCTCCTAAGGTCGAA
 TAAATAAGGAACCAACGATCATTTGGTAAATACCATAGCTTGCCGGAATGTTTCGTTGGCG
 ACCAATCACTGCCTAATAAAAAACGGGCAAAGCCGTAGTTAGCTATGAAAGGTAAGCCAT

TACTAAAAATAAAGAAACAGATTAGCAAAATAGCTACAACAGCTACTGTTGCACTCATGA
 AAAAAATTGCCCTAAAAACTGCTTCTTTGAAGGCTTGTTTTGTCACATCTTGTCTTTCT
 AGTGAAGAAAGTAAGGGAGATACGACACCTCCCTACTTGCCTTCTTTATCTTATTGTACG
 ATGAAACGTCTGCATCTCTTTAGAGATTTATGGAGCAAACATTTTATTTAATCTTGTCCC
 AGGTGGTTAATTTGCCACTAAAAACGTCCGCAAGTTTCAGCCATACTGACTTGGCTTGCCT
 TATTGTCAATTATTGACCACAACAGCAATACCGTCTAAAGCAATAGCATCATGGGTGAGAC
 TCTTACCTTCTTCAGGAGTTAATTCCCTAGAAACCATAACCAATATCAGCGGTTTTCTCCT
 TAACAGCGGTAATACCTGCTGAAGACCCATTAGAGGTAATATCAATCGTAACTTCTGGAT
 TTTCTTTTTTTATAAGCTTCTGCTAATTTTTTCCATTAAAGAAGATACTGAAGTGGAACCTA
 CAACAGACAACCTTGCCTGATAAGTGTGGCTTGTATATTCTGTGGTTTTCGGTTTTAGCTT
 CAATAAATTTATTATCTGTGACCCTTGTGACCTTGTGTTGGAGTGGATAAAGCTGATAA
 AATCTTGACCTAGCTTGGAAAGATTAGAAGACCAAACAATGTTGAAGGGACGTTGAAGAG
 GGTATTCACCATCTAAAACTGTGTCTCGACTAGCCTTGACACCATCAATCTCTAAAGCCT
 TGACAGATTTTCGTTAAAGATCCCAAGGAGATGTAGCCGATAGCATTAGCATTCCCTTGAA
 CTGCTGAGAGAACACCTTCTGTACTATTTTGAATCACAGCTGTTTTGGCAGTGTAGTCAA
 TTTTTTTATCACCGTCTTTTTTGAGAATCCCTGTGATTTCTGTGAAGGCACCCCGTGTTTC
 CAGAGCCATTTTCTCGTGAAATCACCTCAATCGTTCCTGGAGCTGACTGTTTGGAAGCAG
 CTGACTGATTGCCACAGGCAACAAGCCCAAATCCTGATAAGCCAATGGCTGCAAGAGTAA
 GCATTTTTTTTGAATTTTATAATAATCACCTTTATCTCTATGTATTTTTCTGTGTAGGCT
 TACTACATTTATAGTCTAACAAGTCTTTGTAAAGGTTTATCCCTGATTCATGTAAAGATT
 GTGTAAAGAATCAAAAAAGCCACTTTTGAAAAATGGCTGCCCCTAAAAATAG

ORF Predictions:

ORF #	Start	End	Direction	Length
1	68	247	R	60 aa
2	644	1528	R	295 aa

>[SEQ ID NO:153] 3864474-1 ORF translation from 68-247, direction R
 VFMVYNCPKPVYSFLKSAINLMAAIPSIYVGFGLQLLVPWIKTFLGNGMSCPNQLRYY*

Description:

PROBABLE ABC TRANSPORTER PERMEASE PROTEIN (ORF72). - BACILLUS
 SUBTILIS. (BLAST)

>[SEQ ID NO:154] 3864474-2 ORF translation from 644-1528, direction R
 VIIMKFKKMLTLAAIGLSGFGLVACGNQSAASKQSAPGTIEVISRENGSGTRGAFTTEITG
 ILKKDGDKKIDYTAKTAVIQNSTEGVLSAVQGNANAIGYISLGSLSKSVKALEIDGVKAS
 RDTVLDGEYPLQRPFNIVWSSNLSKLGQDFISFIHSKQGGVVDNKFIEAKTETTEYTS
 QHLSGKLSVVGSTSVSSLMKLAAYKKENPEVTIDITSNGSSAGITAVKEKTADIGMVS
 RELTPEEGKSLTHDAIALDGIADVVDNNDNKASQVSMELADVFSKGLTTWDKIK*

Description:

probable hemolysin precursor - Streptococcus agalactiae (strain 74-360)

Assembly ID: 3864510
Assembly Length: 1702bp

```
>[SEQ ID NO:59] 3864510 Strep Assembly -- Assembly id#3864510
CTTTTTTATTTTACACAACAAGTTCATAACGTGTCTTACTGGTGAAGGTTTGACCAGCTTTA
AGAATGACTTGGCCTTTAAGGTCACGTGTGAATGGCATCTGGTAAAGCTTGCGCTTCAAGA
GCAATCCCATTTGTGCTGTAGCATTGGCTGACCTCCTATGATGACACTTTCATCCACAAAG
TTTGCTGTGTAGACCACAAAGCAAGGAGCTTCTGTCTTGAAAAGCAGGAAGCGACCTGAA
TTTTGGTCATAAAGGAATCCAGCATTGTTCATGGCCTGCAGGAAGGGCAAATGGATGATCC
AAACCTGATGCCAGCTGGATTTGCTCATCTTCTTCTGCAAAGATATCCTTCAACAAGGCA
CCATTGTAGATGTGTTTGACCACATCACGGTTGGCTTCTGGAGTTTTGGCAGGAACACCG
TCAGGAGCGATTGAGTAAATGCCCTCTGTGTTTAGTTGGAAGACATGACGGTCAATCGTC
TGCGTGAAATCACCAGACAAGTTGAAATAGCTGTGGTTGGTTGGATTGACCAGCGTATCC
TGATCGGTGCTTACCTTGATAGATCGAATTCATGGAGGCACCAGTTTCTTCCAAGTGATAA
CTGATCGCCAAATCTTGAGATTTCCAGGGAACCTCCTGTCCCATCTGTACGCTCTGTGT
AGAGAGTCAAGCCATGATCGCTTACTTCTTCAACTTCAAACAAGCTGGAATCCCAACCAG
TTGAACCACTGTGATTACAGTTGCTAGCATTATTAACCTCAAGGTCATAGGTCTTACCAT
TGAGCTCAAAGGTCGCACCTGCAATACGACCCGCTACAGGACCTACACTTGCTCCATGCT
TGGGACTATTGCCTACATAACTATCAAAGTCATCAAATCCCAAGATAACATTGGCAAAAT
TTCCAGCCTTGTCAGGTGCGACATAGCGCAAGATAGTCGCACCATAAGTCATAACCTCAA
GTTGGTAGCCACCGTCTGTCTCAAATCGATAGGCCAAGACATCCTCACCTCAACATTTT
CAAATACACGCTCTGTGTATGCTTTTCATTCTGTTCTCCTTTTACTATTTCTCTCAAGCAA
ACAAACCATAGAAAGCGTACTGACAATCTATGGTTTATCTGATAATTTACAAATCCTCTT
GTCAAGAATTCATAAACACTGTCTTACTTTTGATATTCGTGAATTATGACACCTTGACT
ACACGGTTTACTGTACCTGTAGGAGACGGTGTATCTGGTTTATTTTCTACCTTGAGTGAA
GTCAATAGGGCAAAGAGTTGGGCATAAACGATGTAAGGGAAGACACGGTAAATATCATTC
AAGACACCGCCACAACCAAGGGCCACTTCTTTGACATTTTCAAGACCAAAGCTTGATCA
CTCAAAGCACAAACACGACGAGCAATCTGGTCACCAGCAACTTCACGAACCAAGTCCAAG
TCGTACTTACGAGTGTAGTCCGTCGTTGTACCAAAGACCAAACAACACTGTATTGTGCGTTG
ATAAGAGATTTTGGACCGTGACGGAAGCCAACCTGGGCTTTCATACATGGTCGCAACTTGA
CCAGCAGTTAATTCAAAATCTTGAGCTGAGCTTCATGAGCAAGTCCAAAGAAAGGACCA
GCGCCTAGAATAGATGACACGGTTAAAGTCTAAATCAACGAGATCTTTGACATCTTCTGC
CTTGTCTAAAACTTTACGGGCA
```

ORF Predictions:

ORF #	Start	End	Direction	Length
1	1164	1640	R	159 aa

>[SEQ ID NO:155] 3864510-3 ORF translation from 1164-1640, direction R

VSSILGAGPFFGLAHEAQLKILELTAGQVATMYESPVGFRHGPKSLINDNTVVLVFGTTT
DYTRKYDLDLVREVAGDQIARRVLLSDQAFGLENVKEVALGCGGVLNDIYRVFPYIVYA
QLFALLTSLKVENKPDTPSPTGTVNRVVQGVIIHEYQK*

Description:

AGAS PROTEIN. - ESCHERICHIA COLI. (Probable tagatose-6-phosphate
ketose/aldose isomerase)

Assembly ID: 3864526

Assembly Length: 1940bp

>[SEQ ID NO:60] 3864526 Strep Assembly -- Assembly id#3864526
TGCAGGATTTGATTTGGACGACTTTTATTATTACCAGATTCGCCTAGGAATAGAAAAAG
AGCCCAAGAGTTGGACTATGATATCTTGCCTATTTTAATGACCACCCTTTTACCCTAAG
CGAGGAAGTGATTGGGATTCTCTGCATCGGAAAGTTTAGTCGAGCTCAGATTTCTGCCTT
TGAAGAATACCAAAGCCTCTTGTATTTCTAGACAGCGATACACTTTCCTGGGACATAC
CTGTATTATCACGGATTTTACACTGCTATGAAACAGGTTGTCGATTATTTCTCAGTCA
AGGAATGGACCGTATCGGGATTCTAACAGGCCTTGAAGAAACAACAGACCAAGAAGAAAT
CATTCAGGACAAGCGTCTAGAAAACCTCAAAAACCTACAGTCAAGCGAGGGGAATCTATCA
TGATGAACTGGTCTTTCAAGGAAGATTTACTGCCAGTCTGGCTATGACTTAATGAAGGA
GGCCATTCAGAGCTTGGGAGACCAACTTCCGCCAGCATTTTTTCGCAGCCAGCGATAGTTT
AGCTATCGGTGCCCTCCGTGCCCTCCAAGAAGCTGGAATCAGCCTGCCAGATCGCGTCAG
CCTCATTTCTTTAACGACACTAGTCTGACCAAACAGGTCTATCCTCCCCTCTCTAGTAT
TACAGTTTATACTGAAGAAATGGGCCGAGCAGGTATGGATATTCTTAACAAGGAAGTCCT
CCACGGTCGGAAAATCCCTAGCCTGACCATGCTGGGAACCAGACTGACATTAAGAGAAAG
TACCCTAAATCAAGAATAGGATAACATAAAAAACGAATAGAGTTCTAAACTCCTATTTCG
TTTTTTATTTCGATTACAATCATAGACTTAATGGTCTTACGTTTCATCCATATCTTTGTAGG
CTTGGTTCGATATCTTCCAGTTTATAACTTGAAGTAAAGACGCGACCTGGATTGATATCAC
CATCAAGGACGGCTTTTAGTAAAAATTGCTTATCGTATGTTGTAGCAGAAGCTGCCCCAC
CTGCTACAGAGATATTTTGCATAAATGTGCAACCAAGAGCACGATTATTATAGTGTGGGA
CTCCTACAAAGCCCATAACGCCCTCCATTATGAAGAACACCTAGCGCCTGTTCTATAGCAG
CCTCCGTACCAACACATTCAAGTGCTGCGTCTGCTCCTCCGCCGAGGATTTACGCACCT
TGGTAATTCCTTCTTGACCAGTTCTGCAACAACAGCTGTCGCACCTGACTCCATAGCCA
TCTTTTGACGGTCTTCATGACGGCTCATAAGGATAATTTGTGATGCTCCACGCATCTTAG
CCGCGATGACAGCACATTGACCAACAGCCCCATCACCGATAACAACAACCTTGTCCTT
TTTGAACATTTGCAACACGCGCCGCATGATAGCCTGTCGGCATGACATCTGCAAGAGTCA
AAAGGGACTTGAGCATCCCTTCTGTATAGTCAGAAGGTTGACCAGGGATTTTAACCAGCG
CCCAGTTTGCATAGTGGAAGCGAATATATTCTGCCTGAAAATCACCCCCCAAATTATTGC
CAATATGATTGTGCAAGAACCGTCAAATCCAGCAAGACAGGCATCACACTCACCACATC
CATGTGTAAAAGGGACAATCACAAAATCACCTGGTTTTCACCGTCGTAATGGCTTCCCCAG
CTTCTTCAACAATCCCAATCGCTTCGTGTCCACTTATTTTTTGTGTCCAACCTTTCGTTTT
CCNTGGATTACGGTACCTCCATAAATTTGAACCACAAACGCACGCACGAACCACACGAAT
AATCACATCATCCGCTTCTATTATTGCGGACGTTCAATGCTAGCAAGTCCAACCTGACC

TGCCTTTGTATATACTGCTGATTTTCATTTAAAATTTTCCTTCCTTATAAAAGTTTAATTTT
GAGATTTAAACGATTTAAAG

ORF Predictions:

ORF #	Start	End	Direction	Length
1	845	1660	R	272 aa

>[SEQ ID NO:156] 3864526-2 ORF translation from 845-1660, direction R
VKPGDFVIVPFTTHGCGECDACLAGFDGSCDNHIGNNLGGDFQAEYIRFHYANWALVKIPG
QPSDYTEGMLKSLTLADVMPTGYHAARVANVQKGDKVVVIGDGAVGQCAVIAAKMRGAS
QIILMSRHEDRQKMAMESGATAVVAERGQEGITKVREILGGGADAALCECVGTEAAIEQAL
GVLHNGGRMGFVGVPHYNNRALGSTFMQNI SVAGGAASATTYDKQFLLKAVLDGDINPGR
VFTSSYKLEDIDQAYKDMDERKTIKSMIVIE*

Description:

ALCOHOL DEHYDROGENASE (EC 1.1.1.1). - ALCALIGENES EUTROPHUS.

Assembly ID: 3864548

Assembly Length: 2051bp

>[SEQ ID NO:61] 3864548 Strep Assembly -- Assembly id#3864548
ATCGAATTTTTCTAGCCAGGCTACAGTTTTGGCAAGTAAGGTTTCATCTCAGGCAGTCAA
CTGGGTGAGTGCCTTTATTAGCGGAGCTTCTCAAGTGATTGTTGCCTTGATTATCGTTCC
TTTCATGCTCTTTTATCTCTTGCGTGATGGGAAAGGCTTGCGTAACTATTTGACCCAATT
CATTCCAAGAAAATTGAAGGAACCTGTTGGACAAGTTCTATCAGATGTGAATCAACAGTT
GTCCAACATATGTTTCGAGGGCAAGTGACAGTGGCTATTATTGTAGCAGTAATGTTTATCAT
CTTCTTCAAGATTATTGGTCTACGCTATGCGGTTACGCTGGGGGTTACTGCTGGTATTTT
AAATCTGGTCCCTTATCTTGGTAGCTTTCTAGCCATGCTTCCTGCCCTAGTATTGGGTTT
GATTGCTGGTCCAGTCATGCTTTTGAAAGTAGTGATTGTCTTTATTGTAGAACAACTAT
TGAAGGCCGTTTTGTCTCTCCATTGATTTTGGGAAGTCAATTAAACATCCACCCTATTAA
TGTTCTCTTTGTTTTGTAACTTCAGGATCTATGTTTGGTATCTGGGGAGTTTTACTTGG
TATTCCGGTTTATGCCTCTGCTAAGGTTGTCATTTTCAGCCATTTTCGAATGGTATAAGGT
AGTCAGTGGTCTATATGAATTAGAGGGTGAGGAAGTCAAGAGTGAACAATAGTCAACAGA
TGTTACAGGCTTTGGAGGAGCAAGATTTAACTAAGGCTGAGCATTATTTTCGCCAAAGCTT
TAGAAAATGATTCAAGTGATCTTCTGTATGAGTTGGCAACTTATCTTGAAGGGATTGGTT
TCTATCCTCAGGCCAAGGAAATTTACCTGAAAATTGTAGAAGAATTTCCAGAGGTTTCATC
TTAATCTAGCTGCAATGGCTAGCGAGGATGGTCAAAATAGAAAAAGCCTTTAACTATCTTG
AGGAAATCCAAGCTGACAGTGAAGTGGTATGTCTCGCTCTTTGGCTCTGAAGGCAGACCTA
TACCAGCTGGAAGGTTTGACAGATGTGGCACGTGAGAAATTATTGGAGGCCTTGACCTAC
TCAAAGGATTCTCTCTTGATATTGGGTTTGGCAAAGTTGGATAGTGAGTTGAAAATTAC
CAAGCGGCTATTCAAGCCTATGCCCAGTTAGATAATCGCTCGATTTATGAGCAAACGGGC
ATTTCCACCTATCAACGAATTGGCTTTGCCTATGCTCAGTTAGGGAAATTTGAAACGGCT

ACTGAGTTTTTAGAAAAAGCCCTGGAGTTAGAATACGATGACTTAACAGCTTTTGAGTTG
GCCAGTCTTTATTTTGATCAAGAAGAATATCAAAAAGCCACCCTCTACTTTAAGCAGCTT
GATACCATTTCTCCTGACTTTGAAGGCTATGAGTATGGGTACAGTCAGGCTTTACATAAG
GAACATCAAGTTCAAGAAGCCCTGCGTATCGCTAAGCAAGGATTAGAGAAAAATCCCTTT
GAAACTCGCCTCTTGCTAGCTGCTTCACAATTTTCTTATGAATTGCATGATGCTAGTGGT
GCAGAAAAATTATCTCCTTACTGCAAAAGAAGACGCTGAGGATACAGAAGAAATCTTGCTT
CGTTTAGCCACTATTTATCTGGAGCAGGAGCGTTATGAGGATATTCTAGACTTGCAGAGT
GAGGAGCCAGAAAATCTTTTGACCAAGTGGATGATTGCTCGTTCTTATCAAGAAATGGAC
GATTTGGATACTGCTTATGAGCATTATCAAGAGTTGACAGGAGATTTGAAGGACAATCCA
GAATTTCTGGAACACTATATCTATCTCTTGCGTGAATTGGGACATTTTGAAGAAGCAAAA
GTCCATGCTCACACTTACTTAAACTGGTTCAGATGATGTGCAAATGCAAGAACTGTTT
GAGAGATTGTAAGAAATGTTTAAACATATAGAAGTGTAGTTTATCTCTTTTGATAGCTACG
GTCTTTATTTGTACATGGTAGAATCTTTTACAAAAATACTTGGAATCTTGTTTATTCA
TGCCATAATAG

ORF Predictions:

ORF #	Start	End	Direction	Length
1	687	1055	F	123 aa
2	979	1932	F	318 aa

>[SEQ ID NO:157] 3864548-2 ORF translation from 687-1055, direction F
VRKSRVNSQOMLQALEEQDLTKAEHYFAKALENDSSDLLYELATYLEGIGFYPPQAKEIY
LKIVEEFPEVHLNLAAMASEDGQIEKAFNYLEEIQADSDWYVSLFGSEGRPIIPAGRFDRC
GT*

Description:

unknown

>[SEQ ID NO:158] 3864548-3 ORF translation from 979-1932, direction F
VTGMSRSLALKADLYQLEGLTDVAREKLLEALTYSKDSLLILGLAKLDSELENYQAAIQA
YAQLDNRSIYEQTGISTYQRIQFAYALGKFETATEFLEKALELEYDDLTADELASLYFD
QEEYQKATLYFKQLDTISPDEFEGYEYGYSQLHKEHQVQALRIAKQGLEKNPFETRLLL
AASQFSYELHDASGAENYLLTAKEDAEDTEEILLRLATIYLEQERYEDILDLOSEEPENL
LTKWMIARSYQEMDDLDYAYEHYQELTGDLDKNPEFLEHYIYLLRELGHFEEAKVHAHTY
LKLVPDDVQMQLFERL*

Description:

unknown

Assembly ID: 3864582

Assembly Length: 1318bp

>[SEQ ID NO:62] 3864582 Strep Assembly -- Assembly id#3864582
 CTTTAGCAATCAGTTTATTGGGAGATTGACTGCCACTTCTGTTGGAACCTTGATAATCT
 TTTTACCCTCAAAGCGTTCCATACCAGAAATCTTAACATCAACTGCTAAAATAACTACAT
 CCGCTGCATCAATCTGCTCTTGACTCAATTCAATTTCTACCCCTATTGTCCCCTGAGTCT
 CAACATGAATCACATGTCCAGCTACCTTTGCGGCATTCTCTAATTTTTCCTGTGCAATAT
 AAGTGTGGGCAATTCCCATAGTACAAGCTGCAACACCAACAATTTTCATACGGATACCCT
 CCAAAATTTTTTCTTATTAACAAAAGCTGCAATCACATCATCAGATGTCTGAGCCCGAA
 CTAATTTGGCAACAACCTTCGTCAATTACCAAGTTTTCGAGCAAAGAGTGATAAGGTCTTCA
 AATGCTCCCTAGCAGCTTCTGTATCATCACCAACTGCAAAGAGTACAATTACTTTGACCC
 CTTTCCCATCAATGGTCTCCCAAGGAATCTCATTGTGATTTATAGCTATGACTACCCCG
 CCTTCTCCACAGCAGAACTCTAGCTATGGGGAATAGCAATATAATTCCCAATACCGGTCT
 GTCCTTCTGCCTCTCTCTGATAAAGACCTTCGATAAATTGGTCTCTATCAGACACATAAC
 CCGTCTCAACCAATAGTATGAGCTAATGCCTCAAAAACCTCTTCTTTGCTCTGCATCTGT
 AAATCCGTCTGGATCAGACTCACATTAAGAATATCTTTGATTTCCATATATTATCTCCCG
 TAATTCCTTCTTTTGTTAACTGTTTTAATTGATTTATGAATGATTCATCTGCTAGTCTTCT
 CATCAATGTTTTAATACATGACTTGTCTGTGATACTGCAATGGCCAAACCGATAATAAG
 GTCAACACACTGGATATCCTTCGACCATTCTCTGATAGGTGGTTTTAATCTAGTAATCAC
 TAAGACATGATGTTGAAAGTTTCCTTCACAATGTGGTAGAAGAACACCTTTAGCAACCTC
 TATACTTCCCTGTCTCTCACGGTAATATAGAAGCTCTTCTATTTTTTCTGTATCTTCAGA
 AACAAGAAGGCTGATTTGATTTGCTAATTCTTTGTAGGCTTCTTGACGATTTTGAACAGA
 TATATCCATAAGGACAAGCGAAAGATTATTCATAGTTTATCTCCTGAATTTTGTCTTGAA
 GACGTTGTTTATCACCTCGGTTAGAAAAGCACTAACTAGGACAAACGGGACACTTGCTG
 GTTCCTGCAAAGCTACCGTCGTCACAATGAAATCTAAATCTGGATATAGATTTATCAG

ORF Predictions:

ORF #	Start	End	Direction	Length
1	317	550	R	78 aa

>[SEQ ID NO:159] 3864582-1 ORF translation from 317-550, direction R
 VEKAGVVIAINHNEIPWETIDGKGVKIVLFAVGDDTEAAREHLKTLSLFARKLGNDEVV
 AKLVRAQTSDDVIAAFC*

Description:

Probable phosphotransferase enzyme IIa component

Assembly ID: 3864604

Assembly Length: 2077bp

>[SEQ ID NO:63] 3864604 Strep Assembly -- Assembly id#3864604
 CTAGTCTTGGCTACTGTCTAAGTTGGCTTGTGCATAAGCCTGCCAGATTTTTTGTGGGG
 TTTGGCAAGTGGGTAATTCTTGAATTCTTCTGGTGAAAGCCAACGAACCTTCCCTATCTGA
 AAAATCATGGAAGTCACTCACCTGACCTGCTACAATCTGTACATGCCATTTTCGATGACT

AAAACATGCTGGACTGTATCAAAACAAACATCAAGCCAATCAACATCTAGGTCATAGTC
 CTGCTGGAACTCTCTTCTGGGACTGGGGCCAGAGTTCACACTTTCTTCCGCAACCTGAT
 GAAAGAGGTCAAACCTGCTCTTCTTGCAGAAAGTTATCAACTTCTATAAAGGGGAAATGCC
 AAAACCTGCCAAGAGCTTTTCGCTTTCATTTTTTTCAAGTAAAAATTGTCCTTGAGAAT
 TTTTCACAACCTAAGGCTTTAAGATAAATAGGAACCGGCTTTTTCTTAGGAGATTTAATTG
 GATAACGGTCCATGGTTCCATTCTGATATGCCGCACTAAAGTCCTTGACTGGGCTTTCTT
 CAGGTCTGGGATTTACAGGAGACTCAATATCAGACCCTAAGTCCATCAAGGCTTGATTAA
 AATCACCCGGACGATCTGGATTAATCAAGATCTCCATCATTGCCTGAAAAATTTTTCGAT
 TACTTGGAATCCCAATATCGTGGTTGACTTCAAACAGACGCGCCAAGACCCGCATGACAT
 TACCATCTACAGCTGGCTCAGGCAAGTTAAAAGCAATACTGGAAATGGCTCCTGCTGTGT
 AAGGTCCAATCCCTTTCAAGCTGGAAATTCCTTCATAGGTATTTGGAAATTGGCCACCAA
 AGTCAGTCATAATCTGCTGGGCTGCAGCCTGCATATTGCGAACTCGAGAATAATAACCCA
 AGCCCTCCCAAGCTTTTCAGTAACTCTCCTCAGGCGCAGTTGCCAGACTTTTCGACAGTTG
 GAAACCAGTCCAAAAATCTTTTCGTAGTAAGGGATAACTGTATCCACCCTGGTCTGCTGAA
 GCATGATTTTCAGATACCCAGATGTGATAAGGATTTTTACTTCTCCTCCAAGGCAAAATCTC
 TTTTGTTTTTCATCATACCAAGCGAGAAGTTTTCTCACCGGAAAGAAATGACTTTCTCCTC
 CGGCCACATGACGATACCGTATTCTTTCAAATCCTAACATATCTCTAGTTATAACACAGA
 AGGTTTCACCTGTCTTTGTATCTGATTTATAATATTTTTCAATAGATAGTATATAACTTTT
 CCTATCTACTTATACTCCAATGAAAAATCCAAAGAGCAAACCTAAGAAGCTAGCCGCAGGTT
 GCTCAAAACACTGTTTTTGAGGTTGTGGATAGAAGTACAGAGTCAGTATCATATTACCTA
 CGGCAAGGTGAAGCTGACGTAGTTTGAAAAGATTTTCGAAGAGTATAAATCTTATTGATG
 AACTGCTTGACGTCTGAGAAAAAATGAGCTTGGATATTATTTCCAACTCACTTAAAGTC
 AATTTCAATCCACTAGAACAAGCCTAGTACAGTTCCATCGCTTTCAACATCCATGTTGAG
 AGCTGCTGGACGTTTTTGGAAGACCTGGCATGGTCATAACATCACCAGTTAAGGCAACGAT
 GAAGCCTGCACCTAATTTTGGTACCAATTCACGAATGGTAATTTCAAAGTTTTCTGGTGC
 TCCAAGCGCATTTGGATTGTCTGAGAACTGTATTGAGTTTTAGCCATACAAATTGGCAA
 TTTGTCCCAACCGTTTTGAACGATTTGAGCAATTTGTGTTTGAGCTTTCTTCTCAAAGTT
 CACTTTGCTACCACGATAGATTTTCAGTGACAATTTTTTCAATCTTTTCTTGGACAGAAAG
 GTCATTATCGTACAAACGTTTATAGTTAGCTGGATTTTCAGCAATTGTCTTAACAACCTGT
 TTCGGCAAGTGCTACTCCACCTTCTGCTCCATCAGCCCAGACACTAGCCAATTCAACTGG
 TACATCGATTGAGGCACAGAGTTCTTTTAAGGCTGCAATTTTCAGCTTCTGTATCAGATAC
 AAATTCGTTAATAGATACAAGCTAATGGAATACCGAA

ORF Predictions:

ORF #	Start	End	Direction	Length
1	1	141	R	47 aa
2	1513	1803	R	97 aa

>[SEQ ID NO:160] 3864604-1 ORF translation from 1-141, direction R
 VSDFDHDFSDREVRWLSPEEFKNYPLAKPQQKIWQAYAQANLDSSQD*

Description:

unknown

>[SEQ ID NO:161] 3864604-3 ORF translation from 1513-1803, direction R
VNFEKKAQTQIAQIVQNGWDKLPICMAKTQYSFSDNPNALGAPENFEITIRELVPKLGAG
FIVALTGDMVTMPGLPKRPAALNMDVESDGTVLGLF*

Description:

FORMATE--TETRAHYDROFOLATE LIGASE (EC 6.3.4.3) (FORMYLTETRAHYDROFOLATE
SYNTHETAS E) (FHS) (FTHFS). - CLOSTRIDIUM ACIDI-URICI.

Assembly ID: 3864610

Assembly Length: 1887bp

>[SEQ ID NO:64] 3864610 Strep Assembly -- Assembly id#3864610
CTCAAAACNCTGCTTTGAAGAGATTTTCAAAGAGTACAAGAAGTTAGTTATTAGCGTTC
TTACCGCTTGTAAGTAGATTTCTCATAAAATAGAATCTTTTCCTTTTAGTTGTAAACTA
GTCTGGGAGAGTAGAGAGGTTTGAGATACCTTTCTAGCTTTTGGATTATCATCTAAGAAG
AGTAATTTCCCTTGCAATTAAGAGGGGAGAGACACGAAATGACTATAATGGGTGAC
AATGGGGGAAGGATAGACAAGAGATTTTATCCACATATGAAAAAGGAGGTTAGGAAAG
AGTTATATATCCTATATTATATAAATAATCAATTGCGCAGAAATTTGGTAAGAATTCATG
CGTCAACTCATAAAGAACTACTTAAAAAATTCACAGTATTCATAATTATTTTCGAGGAGA
AAAACAGTGAAAAAAGAAAAAGCTTGCTCTGTCTCTTATCGCTTTTGGCTGACGGCT
TGTTTAGTAGGCTGTGCTAGCTGGATTGATCGTGAGAAATCCATAACGGCTGTTGGCTCA
ACTGCCTTGCAACCTTGTTGAAGTAGCGGCAGATGAATTTGGCACCATCCATGTTGGA
AAAACGGTCAATGTCCAAGGGGAAGTTCTGGTACAGGCTTGTCACAGGTTTCAAGTCTGGG
GCAGTTGATATAGGAACTCAGATGTATTTGCTGAGGAAAAAGACGGAATTGATGCTTCT
GCTCTTGTTGACCACAAGGTCGCGGTAGCTGGCTTGCTCTGATTGTCAATAAGGAGGTT
GATGTTGATAACCTAACGACAGAGCAACTTCGTCAAATCTTCATAGGTGAGGTAACCAAT
TGGAAGAGGTTGGTGGTAAGGACTTACCCATCTCTGTTATCAATCGGGCAGCCGGCTCT
GGCTCTCGTGCTACCTTTGATACTGTCAATTATGGAAGGTCAGTCTGCCATGCAAAGTCAG
GAGCAGGATTCAAATGGAGCGGTAAATCAATCGTATCAAAAAGTCCAGGAGCTATCTCT
TATTTATCTCTTACCTATATAGATGATTCCGTCAAAGCATGAAGTTGAATGGCTATGAC
TTAAGTCCAGAAAATATAAGTAGCAATAATTGGCCCTTGTTGGTCTTATGAGCATATGTAT
ACATTGGGGCAGCCCAATGAGTTGGCTGCAGAAATTTCTCAATTTTGTCTCTCGGATGAG
ACCCAAGAAGGGATTGTCAAAGGATTGAAGTATATTCCGATTAAGGAAATGAAGGTTGAA
AAAGATGCTGCCGGAAGTGTGACAGTGTGGAAGGGAGACAATAATGAATCAAGAAGAAT
TAGCTAAGAAAATGTTGCTTCCATCAAAGAATTCTCGTCTGGAGAAATTAGGAAAAGGTT
TGACCTTTGCCTGTCTTTCTTTGATAGTCATCCTTGTTGGCCATGATTTTGGTTTTCGTAG
CGCAAAAAGGCTTGTCGACCTTCTTTGTCAATGGTGTGAATATCTTTGACTTTCTTTTGG
GAGGAACTTGGAATCCTTCTAGTAAAGAATTTGGTGGCCCTTCCTATGATTTTGGGTTCTCT
TTATCGTTACCATTCTCTCAGCCCTTATCGCAACACCCTTTGCTATTGGTGCAGCAGTTT
TTATGACCGAAGTATCACCAAAAGGGGCGAAGATTTTGCAACCAGCTATTGAATCCTGG
TTGGGATTCTTCAGTAGTGACGGATTTATTGGCTTGCAAGTCGTCGTTCCCTTTGTTC

GCAGTGTCTTTGGTGGGACTGGTTTTGGGATTTTGTTCAGGGATTTCCGTCCTCTTTGTCA
 TGATTTTGGCCGACCGTAACCTTTATGACAACGGATAGCTTGCCTGCGGTTCCCTCCNTTAT
 TATCGTGAAGCCAGTTTCGCTATGGGA

ORF Predictions:

ORF #	Start	End	Direction	Length
1	427	1305	F	293 aa

>[SEQ ID NO:162] 3864610-1 ORF translation from 427-1305, direction F
 VKKRKKLALSLIAFWLTACLVGCSWIDRGESITAVGSTALQPLVEVAADFEFGTIHVGKT
 VNVQGGSSGTGLSQVQSGAVDIGNSDVFAEEKDGIDASALVDHKVAVAGLALIVNKEVDV
 DNLTTTEQLRQIFIGEVTNWKEVGKDLPISVINRAAGSGSRATFDTVIMEGQSAMQSSEQ
 DSNQAVKSIVSKSPGAISYLSLTYIDDSVKSMKLNQYDLSPENISSNNWPLWSYEHMYTL
 GQPNELAAEFLNLFVLSDETQEGIVKGLKYIPIKEMKVEKDAAGTVTVLEGRQ*

Description:

PROBABLE ABC TRANSPORTER BINDING PROTEIN PRECURSOR (ORF108). -
 BACILLUS SUBTILIS. (BLAST)

Assembly ID: 3864716

Assembly Length: 405bp

>[SEQ ID NO:65] 3864716 Strep Assembly -- Assembly id#3864716
 CTGAGGAATCAAAAGTTGAACCACCAGTAGAACCAAGCATAAGTCCCAGAACAAACCCGTGC
 AACCTACACAAGCTGAGCAACCAAGTACACCAAAAGAATCATCACAACAAGAAAATCCTA
 AAGAAGATAGGGGAGCGGAAGAGACTCCGAAACAAGAAGATGAACAGCCAGCAGAAGCCC
 AAGAAATCAAGGTTGAAGAACCAGTAGAATCTATAGAGGAGACTGTCATTCAACCTGTTG
 AACAAACCAAAAGTGGAAACGCCTGCTGTTTAATAACTAACGGAACCTACAGAGGAACCTA
 AAGTTGAAGTAAGTATATCCCTCACTACTCGCTATGAGGAAGACCTTACTTACGAAC
 ACGGAACGCGTTGAAGTTGTTAAGGAAGGTTATAATTGGCAGTAT

ORF Predictions:

ORF #	Start	End	Direction	Length
1	57	272	F	72 aa

>[SEQ ID NO:163] 3864716-1 ORF translation from 57-272, direction F
 VQPTQAEQSPSTPKESSQQENPKEDRGAEETPKQEDEQPAEAQEIKVEEPVESIEETVIQP
 VEQPKVETPAV*

Description:

unknown

Assembly ID: 3864718
 Assembly Length: 1542bp

>[SEQ ID NO:66] 3864718 Strep Assembly -- Assembly id#3864718
 CTATGGGATTGGTAGTTCTTCCTAGTGCAGGGGCTGTAGACCCAGTTGCGACCCTAGCGC
 TGGACTAGTCGAGAGGGTGTGTTGTTGAAAATGGATGGCTATCGCTATGTTGGTTATCTATC
 AGGTGACATCCTCAAAACGCTTGGCTTGGACACTGTTTTAGAAAGAAACCTCAGCAAAACC
 TGGAGAGGTGACTGTAGTCGAAGTTGAGACTCCTCAATCAACAACAAATCAGGAGCAAGC
 TAGGACAGAAAACCAAGTAGTAGAGACAGAGGAAGCTCCAAAAGAAGAAGCACCTAAAAC
 AGAAGAAAGTCCAAAGGAAGAACCAAAATCGGAGGTAACCTACTGACGACACCCCTTCC
 TAAAGTAGAAGAGGGGAAAGAAGATTCAGCAGAACCATCTCCAGTTGAAGAAGTAGGTGG
 AGAAGTTGAGTCAAAACCAGAGGAAAAAGTAGCAGTTAAGCCAGAAAGTCAACCATCAGA
 CAAACCAGCTGAGGAATCAAAAGTTGAACCACCAGTAGAACAAGCAAAAGTCCCAGAACAA
 ACCCGTGCAACCTACACAAGCTGAGCAACCAAGTACACCAAAAGAATCATCACAACAAGA
 AAATCCTAAAGAAGATAGGGGAGCGGAAGAGACACCGAAACAAGAAGATGAACAGCCAGC
 AGAAGCCCAAGAAATCAAGGTTGAAGAACCAGTAGAATCAAAAGAGGAGACTGTTAATCA
 ACCTGTTGAACAACCAAAAGTGGAAACGCCTGCTGTAGAAAAACAAACGGAACCAACAGA
 GGAACCAAAAGTTGAAGTAACAAGTATTCCCCAACTACTCGCTATGAGGAAGACCTTAC
 TAAGGAACACGGAACGCGTGAAGTTGTTAAGGAAGGTAAGAATGGCAGTAGAACAGTTAC
 TACTCCATATATCTTGAATGCGACAGATGGTACGACTACAGAAGGCACTTCGACAACCTGA
 TGAAGCTGAGATGGAGAAAGAGGTTGTTTCGTGTTGGCACGAAACCCAAAGAAAAATTAGC
 TCCAGTCTTAAGTTTGACAAGTGTTACAGATAATGCAATGTTGCGTAGTGCGAGACTTAC
 TTATCATTTGGAAAATACAGATAGTGTGATGTGAAAAAAATTCATGCTGAAATTAAAAA
 TGGCGATAAGGTTGTCAAACTATTGACTTATCTAAAGAGAGATTATCAGATGCTGTTGA
 CGGTCTTGAACCTTTATAAAGATTATAAGATTGTGACGAGTATGACCTATGATAGAGGTAA
 TGGTGAAGAAACCTCTACGTTGGAAGAACTCCACTACGATTAGACCTCAAGAAGGTTGA
 ATTGAAAAACATCGGCTCTACTAATCTCGTCAAAGTAAATGAGGATGGTACTGAGGTGGC
 AAGTGACTTCTTAACAAGTAAACCTGTGGATGTGCAGAATTACTACCTCAAAGTAACTTC
 CCGTGATAATAAAGTTGTTTCCCCTCCCAGTTGAAAAAATTGAAGAGGTGACTGAGGAAG
 GTCCACCACTTTACAAAGTCCCTGCTAAGGCCCTAATTTGAT

ORF Predictions:

ORF #	Start	End	Direction	Length
1	77	1474	F	466 aa

>[SEQ ID NO:164] 3864718-1 ORF translation from 77-1474, direction F
 VLLKMDGYRYVGYLSGDILKTLGLDVLLEETSAKPGEVTVVEVETPQSTTNQEQARTENQ
 VVETEEAPKEEAPKTEESPKEEPKSEVKPTDDTLPKVEEGKEDSAEPSVVEEVGGEVESK
 PEEKVAVKPESQPSDKPAEESKVEPPVEQAKVPEQPVQPTQAEQSPSTPKESSQQENPKED
 RGAEETPKQEDEQPAEAQEIKVEEPVESKEETVNQPVQPKVETPAVEKQTEPTEEPKVE
 VTSIPQTTRYEEDLTKEHGTREVVKEGKNGSRVTVTPYILNATDGTTEGTSTTDEAEME

KEVVRVGTPKEKLPVLSLTSVTDNAMLRSARLTYHLENTDSVDVKKIHAIEIKNGDKVV
 KTIDLKERLSDAVDGLELYKDYKIVTSMTYDRNGEETSTLEETPLRLDLKKVELKNIG
 STNLVKVNEDGTEVASDFLTSPVDVQNYLKVTSRDNKVVSPPS*

Description:

unknown

Assembly ID: 3864802

Assembly Length: 1321bp

>[SEQ ID NO:67] 3864802 Strep Assembly -- Assembly id#3864802
 ATCGAATTACTTCAACTCCAACCTTTACTCTCAATAAAAATCAAATGTAAAAAGAGGAGCT
 AAATTTATCTTTTTCTCCTCCTTCATCGTTCTTACTTTTGACCATAATAAGCATTGTC
 CATGTTTACGTTGGTAGTGTTTTCTAGTATGTACTGGGGAGCAGGTTCAACTCTTGAT
 TGATTTGTTCTGTAAAGCGATTCATCTTTGATACTTCCTCTAGTACGACAGAGTGATAAA
 CAGCATTCCTCGGATTTTTGCCCCAGGTGAATGGACCGTGATTGCGTACAACAATTCCTG
 GTACTTCAACCGGGTTAAGTCCGCGATGTTCAAACCTTCTACGATAACCAGGCCAGTAT
 CTTTTTCATAGGCCACTTCTACTTCGTCCTTGGTCAAACCTACGGGCGCAAGGGATTGAAC
 CGTAGAAATAATCTGCATGGGTGTTCCGTAGAAAGGAATATCACGACCTGCCTGAGCCC
 AAGCAACAGCTTCTGTGCAATGGGTGTGAACCACACTACCAATTTCTGACCAAGCCTTAT
 ATAATTGCACATGAGTTGGGAAGTCGGAAGATGGTCTTAAATCCCCTTATAGGATCTTAC
 CATCTAGATCAGTCACTACCATGTTTTTCAGGTGTCAATTTCGTCATAATCCACGCCTGATG
 GTTTGATAACAATGACACCGAGTTCGCGATTGACTTCAGATACATTCCCCCAGGTAAATT
 TGACAAGTCCATGTTTTGGCAATGATTGATTGGCATCACAGACTCGTTTACGCATAGCAT
 TGATTACTTGATTTCATCTTACATCAAACCTGCTTTCTTAATGAGTGGATAGAGAAAAGCT
 TCGCCTCTTGAATGGCTGCGCGTGTTCCTTCTACTGTTTCACAATTTTCAGACCACATT
 TCGATTAGGAAAGGTCCATTATAATTGGTTTCCTTTAAATATCGAAAGCTTCTTCCCAT
 TTGACACAACCTTGCCCCAAAGGTACATCTCGGAAGTGGCCCTTTGAACTTTCTGTCACT
 GCATAAGTATCCTTGAGATGGAGAGTTGCGATGGCATGATGACCAAGATAAACTCACTA
 TAGATATCATTATGCCATGCAGACACATTACCAATATCTGGATATACAAAGAGGAAGGGA
 GAGTCAATCTCTTTTTCTATAGCCAAATATTTTTTCGATGCTATTGATGAAAGGATCATCC
 ATAATTTCAATAGCAAGTACCACCTGAGCTTCTTCAGCCCAGTCACAGGCTTTTCTCAA
 TTTTGTGATAAACGTTGGCGTGTCTGGGGTGACTTTTCCTCATAGTAAACATCGTAACCA
 G

ORF Predictions:

ORF #	Start	End	Direction	Length
1	92	550	R	153 aa

>[SEQ ID NO:165] 3864802-1 ORF translation from 92-550, direction R
 VQLYKAWSEIGSVVHSTHSTEAVAWAQAGRDIPFYGTTHADYFYGSIPCARS�TKDEVEVA
 YEKDTGLVIVEEFEHRGLNPVEVPGIVVRNHGPFWTGKNPENAVYHSVVLEEVSKMNRFT

100

EQINPRVEPAPQYILEKHYQRKHGPNAYYGQK*

Description:

L-RIBULOSE-5-PHOSPHATE 4-EPIMERASE (EC 5.1.3.4). - ESCHERICHIA COLI.

Assembly ID: 3864854

Assembly Length: 1265bp

```
>[SEQ ID NO:68] 3864854 Strep Assembly -- Assembly id#3864854
TTTTTCTGTTTTTCGGAGCAAACCTGGGCTCCAGCCGGTTTTGGCCTTCTTTCTTAGCTA
CAGCTGGTTTAGCTGGCTCAGATTTTTTCGGCTTTCTTTTCTGCACTTACTTTTGGTGCTG
CAGGTTTTGCTTCTACTTTTCGGAGCAGCTGCAGGCTTAAAGCTGGCAGCAATTTTGCAG
CGACAGCTTCTTCCACACTTGATGAGTGGCTTTTCACATCCAAGCCCACTCTTTTGCAC
GCGCTACAACCTTCTTTACTTTCTTTTCCAAGTTCTTTTGCGATTTTCGTACAATCTTTTCT
TAGACAAATCATGTCTCTCTTCTATTCCATAAGAGACCTCATTTTCTTTGTAAATCCA
GCATCTGTTACAGCCAAAACCTTTCTCGATTTCCCGACTGCTATGATTAATTCCAGTGTT
GAAAACACGGTTACAATTTCTACTTGATAATAATGACTTTTATCTTGAATCTTCTTGGTC
AGATTGGGTCCAGCATCATGAGCTAGAAAGACCAACTTGGCCTTGCCGTCTTGAATGGCC
TTGACCACCAATTCTTCACCCGATATGATGCGCCCTGCTCGCTGAGCAAGCCCCAAGAGA
TTACTTATCTTTTGCTTATTCAAGTCCCAACTCTCTTCTTTTCACTTTGTGATCCACATA
AGCGATCAACTCGTCATAAAAGCTTTCTTCCACTTCCATGCTAAAGCTGCGGTTAAAGAC
CTTCTTCTTTTTTCGCCTCTAGGGCTTCTGCATTGTCTAGTTTGATATAAGCGCCGCGGCC
ATTGGCCTTGCCCGTAGGATCAATAAAGACTTGTCTTCTTCTTGTGACAATGCGGAG
CAAATCACGCTTATCAATCACTTCGTTAGACACAACAGACTTGCGCAAAGGGATTTTCT
TGTTTTCATCTTTCCCTCCTCTAGCAGCTTTTATTCTTCTACAGTATCGTTTTTCTACTTC
CAACTCTACTGAAGCAGCGTCTTCCATGGCTTCAAATTTCGCTAGCAGACTTGATATCGAT
ACGGTAACCAGTCAAGTGAGCCGCCAAGCGCACGTTTTGTCCACGACGACCAATGGCAAG
AGAAAGCTTGTTATCTGGAACAACCAAGGCACGTTTGCTGTCGTTTTCATCAAAGAT
AACTTGGTCAACCTCAGCAGGAGCGATGGCATTGTAGATAAATTCAGCTGGATCTGCTAC
CCACTCGATAACATCGATATTTTCTTCGATTGGTACCATGCGGTCATTTTTAGCATCGTA
ACGAG
```

ORF Predictions:

ORF #	Start	End	Direction	Length
1	324	548	R	75 aa

```
>[SEQ ID NO:166] 3864854-1 ORF translation from 324-548, direction R
VVKAIQDGKAKLVFLAHDAGPNLTKKIQDKSHYYQVEIVTVFSTLELIIVGKSRKVLAV
TDAGFTTKMRSIME*
```

Description:

PROBABLE 11.1 KD RIBOSOMAL PROTEIN IN NUSA-INF8 INTERGENIC REGION
(ORF4). - BACILLUS SUBTILIS.

Assembly ID: 3864862

Assembly Length: 1305bp

```
>[SEQ ID NO:69] 3864862 Strep Assembly -- Assembly id#3864862
ATAAACCAAAGGAAGCTGAGCTCTTTAGTCCCAGCTTCTTTTTATATATAAAATTTTACC
CGTGAAAGACAGGGCCTTAGCAGACTTCTTTTTTACTTCGTTACCCTTGCTTTTTCTT
TGTATGTTTGGGCGTTGGCAGTTGGTTATACATAGCTAAAATCAGGTCTTATAGAAACAT
CTTATTATCAAGTTCTTCCACTCAAATCATTTCTTTGGCACCTTTGTATGGAAACTCAA
AGAAGATTGGTCAATCTTATCTAAGACTGCTTGCACGGGTTTAACTAAAAGCGATCGTCA
TAAATGCCGCCAATAATCTTGCCGCGGAAGTAAAGAATATACTCCCCATCATGGAACGG
TAAGTCACATCATCTAATCCTGATAATTGTTCCAAAACAAATTCCAAATAGTTCTTACTT
GATGCCATTTCTAATCTTCTAGGCTCTGTTCAACGATAACAACCGTATAGAGTTCTTGCT
TAACCTCGCATCCAATTGATTTAAAGCCCTGCTTTTCCCAAAAATGCTGAGATTGCGGAT
TTCCCTTAACATAAGCCAAACGTGCCTTTCGAAAGTTCTTAGCAAAAATAAGCTAGTGCTT
CTGTCACAATATGACTACCAATCCCTTTTCTCTGATAGGCTTGATCAACCATAAAACAAAC
CAATAAAAACAGTCTCCTCATCAGGATATGCATAGACAAAATCCATAACAGCCACAAGGT
CAAATCCATTCCAAAATCCAACAAAAAACTTATCAGCCTTAGCTTTACCTTCAGGTAGAC
AAAGCATGTCCTCTTTTACAGTTGCAAAATTTGGCTCTGGTGGACAATGCTGAAAATACA
GAGGATTACTTTTCATATAAAGATAAAATACTTGGAATATCCTTTTTCAGTTAGTATCCTAC
AACTGTAATACTTAGATAGTTGGTCAATCATCTTTTCAAATTCGATACTTTCTTGCGCC
TGTGATTATGACACAGGAAGATGCACTGATCGTCATCAGCCACATAAAAGTTCTTTCCAT
CGTGCCTAATCGTTGTCTCAAACCTTTGGATAAAACCTTTAGCCTATACAACTGGATTTT
CCTCTCTCAAAAGTATATTTCTTTTGCAGGCGAACTTCCTCAAAATCAGTCGTGTGCAACT
TCAGTAGAATATTCATAGGCTCGGATAATCTGAGCGACAACAGGATGGCGAACCACATCC
TTGGCTGAAAAATGAACAAAGTCAATCTGATGGATGTTCTTGAGTTTCTCTTGAGCATCA
ATCAAACCGGACTTGACATTACGTGGCAGGTCAATCTGACTAATA
```

ORF Predictions:

ORF #	Start	End	Direction	Length
1	431	1003	R	191 aa

```
>[SEQ ID NO:167] 3864862-1 ORF translation from 431-1003, direction R
VADDDQCIFLCHNHRAQESIEFEKMQDQLSKYYSCRILTEKDIPSILSLYESNPLYFQHC
PPEPNFATVKEDMLCLPEGKAKADKFFVGFWNGFDLVAVMDVYAYPDEETVFIGLFMVD
QAYQRKGIGSHIVTEALAYFAKNFRKARLAYVKGNPQSQHFWEKQGFKSIGCEVKQELYT
VVIVEQSLED*
```

Description:

unknown

Assembly ID: 3864888

Assembly Length: 1742bp

```
>[SEQ ID NO:70] 3864888 Strep Assembly -- Assembly id#3864888
CTAATCTCCTTAAAACGTGATCTTTTCAAGAATATTTTATCTAAACAATCCAGCAAGTC
TTGGTAAGAATAGACTTCGTAAGTCGGCTGGGCTTGTGTGTGATTTTCGAGGTGATGAGG
ATTATACCAGATAGTGTCAATCCCCGCATTATTGCCACCTTGAATGTCGGCGGTTAGAGA
ATCTCCAATCATCAGCGTCTTTTCTTTACTAAATCCAGCAATTTGCTGGCCAATCTTTTC
ATAAAAAAGAGCATCCGGCTTTTGAGTTTGCAACTGTTCTGAGATAAAGACTTGATTGAA
ATAAGGTGCTAGACCAGATTGAGCCAAACGTCCTGTCTGAATGGCAGTAATGCCATTTGT
CGCAGCATACAAGTTATAATCACGCTCAATGAGGCTGTCCAAGAGATCATGAGCGCCCGA
TAGTGTGTTGTCCCTGCTGGGCGAGGTAAAATTGGTAACGCTGGGCAAGAAAACCTACCGTC
TTTTTCCTGTCCAAAATGAGCAAATAAACGAGAAAAGCGCGTGTTAACCAGCTCTTGTTT
ACTGATTTTCTTCAGCTCCAAGTCTTTCCAGAGAGCCTTGTTTCATAGGAACGTAATAATC
TTTATAAGCCGGAATATCCGCAACTCCTTCTTCTTTTAGAAGTGGAGTCAAAGCCACATC
CTCAGCAGCATCAAATCAAGAAGAGTGTGGTCGAGGTGGAAGAGTACAAATTTGTAGAA
CAATTTGAGGTTTTCTTTCTGAAAATTCATTAAGAACATTATATCATAAAGCACCTCAT
ACAATTAATAATTTAATCACTTAAAAAAATTTCGAACACTTCTATACAACTGACAGCT
CAAATCTTTTCAGAATAGAACAAATACTAATATCGAACACCCCGTCTTCATAAATACATAT
GTAATTCTAGGCCCTAGAATTCCTATAAACTAAATGCTTTTCATACTCTTCCAAGTAATTGA
TTGCCTTAAATTTTAATTTTGAAGGTTTCTAAAGCTAGAATAGCCCCATCACAATCAGT
TTTGATTGATTCACAATTTAGAAACACTATAGTTTCACTCCTGTAAAATAAAAAGGAAC
TGCATAAAGCAATCCCTTTCTGATTTTGAATCATTTACTTAACATTTTATAGTTGAGAT
AATCAATAGCTTATCTATAAAAAGAGTTATAGTAAAATTCCTTATTTATTGATTCCAAGC
TCCGCTAACTGTATTTGAATAACTGACAGTTCTGCACCAGCCTGAAAAAGAGCAGCTGCA
TTATAGGCACCTTCTACAATTGGAACCCTGTTGATGATGATACTTTTATCACTGAAATCA
GTCACCATTTTTTAAGTTCATTTTAGCAGAACCTAGGTCAAAAAGGCAAGTAAAGTATCT
GCTGGATTTTTCGGAAACAACCCTATCTACTTGATCAAACTCGTTCCAATTCCTCCGCCC
TCGGTTCCTCCTACATAAGTAATCGGAACATCTTTAGCTACTTTACTAATCAGTTCAACA
ACACCTTCTGCAATGTGTTTGAATGTGAAACGATAACAAGACCAATACCAATACTTTCC
ATCAAACCACTCCAGTTTCTAAAATAGCAGTAAAGAGTAATCCTGATGAGAATGATCCAG
GATCAATATGTCCAAGAAACCACATGCTCCTAAGACAAGAGCTAACAGACTGGCCATCAA
TAATAGTATTGTTCTTTTTTTTCATCATTACTCCTTAAGTAGTGTAACTGATTAATTGC
AT
```

ORF Predictions:

ORF #	Start	End	Direction	Length
1	10	657	R	216 aa

```
>[SEQ ID NO:168] 3864888-1 ORF translation from 10-657, direction R
VALTPLLKEEGVADIPAYKDYYVPMNKALWKDLELKKISKQELVNTRFSRLFAHFGQEKD
```

GSFLAQRYQFYLAQQGQTLGSAHDLLDSLIERDYNLYAATNGITAIQTGRLAQSGGLAPYF
 NQVFISEQLQTQKPDALFYEEKIGQQIAGFSKEKTLMIGDSLTDIQQGNNAGIDTIWYNP
 HHLENHTQAQPTYEVYSYQDLLDCLDKNILEKITF*

Description:

unknown

Assembly ID: 3864898

Assembly Length: 1136bp

>[SEQ ID NO:71] 3864898 Strep Assembly -- Assembly id#3864898
 GTGGAATGCGGGGACGCCTTGTCTAATTTTGGATCAAGCCCTGAGTTTGACACAGGGAAA
 TGAGCTGGACGGACTGCTATCTCTGAAGAAATTACTGGCACCATTAGCCTATCAGCCTTG
 GATGATTATGTGGCGGCCTTGTCTCAACAGGATGTTCCCAAAGCTTTGTCTTGCTTGAAT
 CTTCTTTTTTGACAATGGTAAGAGCATGACTCGTTTTGTGACCGATCTTTTGCACATTTTA
 AGAGACTTGTTAATTGTTCAAAACAGGGGGAGAAAATACTCATCATAGTTCAGTCTTTGTA
 GAAAATTTGGCACTTCCTCAAAAAAATCTGTTTGAAATGATTTCGCTTAGCAACAGTGAAT
 TTAGCAGATATTAAGTCTAGTTTGCAGCCCAAGATTTATGCTGAAATGATGACCGTCCGT
 TTGGCGGAAATCAAGCCCGAACCAGCTCTATCAGGAGCGGTTGAAAATCGAATTGCTACG
 CTGAGACAGGAAGTTGCCCCGTCTCAACAAGAGCTTCTAATGCAGGTGCGGTTCTCTAAA
 CAAGTTGCACCAGCTCCTAGTCGACCAGCTACGGGCAAAACAGTCTATCGTGTCTGATCGC
 AATAAAGTGCAATCTATCTTACAAGAGGCCGTCGAAAATCCTGATTTAGCACGTCAAAAT
 CTAATTCGTTTGCAGAATGCCTGGGGAGAGGTAATTGAAAGTCTAGGTGGGCCGGACAAG
 GCTCTGCTAGTTGGTTCTCAACCGGTTGCTGCCAATGAACACCATGCTATTCTTGCTTTT
 GAGTCTAACTTCAATGCTGGTCAAACATGAAACGAGACAATCTCAATACCATGTTTGGT
 AATATCCTCAGTCAGGCGGCAGGTTTTTTCACCTGAGATTTTAGCTATTTCCATGGAGGAA
 TGGAAGAAGTTTCGCGCAGCCTTTTCAGCCAAAGCCAAATCTTCTCAAACGAAAAAGAA
 GTAGAAGAAAGCCTGATTCCAGAAGGATTTGAATTTTTGGCTGATAAAGTGAAGGTAGAG
 GAAGACTAAAGAAAGATTTTCATGATACAATAAGTTTATGAATAACAACAATTTATTATT
 ATGGCGCTATTTACAGCTGCTGAGACCTATTTTTTCAATGAAGCCTGGATGACTGG

ORF Predictions:

ORF #	Start	End	Direction	Length
1	130	1029	F	300 aa

>[SEQ ID NO:169] 3864898-1 ORF translation from 130-1029, direction F
 VAALSQQDVPKALSCLNLLFDNGKSMTRFVTDLLHYLRDLLIVQTGGENTHHSSVFVENL
 ALPQKNLFEMIRLATVNLADIKSSLQPKIYAEMMTVRLAEIKPEPALSGAVENRIATLRQ
 EVARLKQELSNAGAVPKQVAPAPSRPATGKTVYRVDRNKVQSILQEAVENPDLARQNLIR
 LQNAWGEVIESLGGPDKALLVGSQPVAANEHHAILAFESNFNAGQTMKRDNLNTMFGNIL
 SQAAGFSPEILAI SMEEWKEVRAAFSAKAKSSQTEKEVEESLIPEGFEFLADKVKVEED*

Description:

unknown

Assembly ID: 3864938

Assembly Length: 1670bp

```
>[SEQ ID NO:72] 3864938 Strep Assembly -- Assembly id#3864938
CTGTCTCTGAAACAGTCACATCAAGTGCCTCTGAACAANC GCCCNCCTAGGTNGACGGT
ATCGATAAGCTCGATCTGTGATTTTCAGAGAAGAAATCAAGTGTGTAACAGAAGTAAGAT
GTAATTGTATGTAAAGGAGACGTCATGTTAAATAGTATTGTAACCATTATTTGTATTGCC
CTTATCGCGTTTTATCTTGT TTTTGGTTTTTCAAAAAGCCTGAAAAATCTGGACAAAAGCC
CAGCAAAAAAACGGATACCAAGAGATTCGAGTGGAAGTCATGGGAGGCTATACTCCTGAG
TTGATTGTCCTCAAGAAATCAGTGCCAGCCCGCATTGTCTTTGACCGCAAGGATCCTTCA
CCATGTCTGGATCAAATTGTTTTTCCAGATTTTGGTGTACATGCGAACCTGCCAATGGGG
GAAGAGTATGTAGTGGAATCACGCCTGAACAGGCTGGAGAGTTTGGCTTTGCTTGTGGT
ATGAACATGATGCACGGCAAGATGATTGTAGAGTAGGTGGAGACTATGACAGAAATTGTG
AAAGCAAGCTTAGAAAAATGGCATTCAAAAAATCCGTATCCGAGCTGAAAAGGCTATCAT
CCAGCCCATATCCAGCTTCAAAGGGAATTCAGCTGAGATTACCTTTCATTCGTGCTAC
TCCTTCAAACCTGTTATAAGGGAAATTCTGTTTGAAGAAGAAGGTATCTTGGAAGCAATCG
GCGTAGATGAGGAGAAAGTCATTTCGTTTTACACCTCAAGAATTAGGGAGACATGAATTTT
CTTGTGGCATGAAGATGCAAAAGGGAAGCTATATAGTCGTTGAGAAGACTCGAAAATCTC
TATCTCTCCTGCAAACGTTTTTGGATTACTAGTATCTTTACTGTGCCTCTTGTGATTCTC
ATGATTGGGATGTTGGCAGGTAGCATTAGTCATCAAGTCATGCATTGGGGAACCTTTTTTA
GCAACAACGCCTATTATGTTAGTTGCGGGTAAGCCATATATCCAGAGTGCTTGGGCCAGT
TTTAAAAAGCACAAATGCCAACATGGATACCTTGGTTGCGCTGGGAACCTCTAGTGGCTTAT
TTCTATAGCCTAGTTGCTCTCTTTGCTGGTCTCCCTGTTTACTTCGAAAGTGCTGGATTT
ATCCTCTTTTTTCGTTCTTTTTGGGAGCAGTTTTTTGAGGAAAAAATGAGGAAAAATACGTCC
CAAGCTGTGGAGAAATTACTGGACTTGCAAGCTAAAACCGCAGAAGTCTTGAGTGATGAT
AGTTATGTCCAAGTTCCTTTGGAACAAGTCAAGGTACGCGACCTTGATTCCAGTGCGTCC
CGGTGAAAAGATTGCTGTTGATGGTGTCTAGTAGAAGGTGTCTCTAGTATTGACGAATC
CATGGTGACAGGTGAGAGTCTGCCTGTGGACAAGACAGTTGGAGATACTGTCATTGGCTC
AACCATCAATCATAGTGGAACGCTTGTCTTTAGAGCAGAAAAAGTTGGCTCAGAGACTGT
TTTGGCTCAGATTGTAGATTTTGTGAAGAAAGCTCAGACAAGTCGTGCGCCGATTCAGGA
CTTGACGGATAAGATTTCAAGGATTTTTGTCCCAGTAGTTGTCATTTTAGGAATCATGAC
CTTTTGGGTTTGGTTCGTCTTGCTCAGGGATAGTGTGGTCGTGCTTGGAG
```

ORF Predictions:

ORF #	Start	End	Direction	Length
1	883	1326	F	148 aa

```
>[SEQ ID NO:170] 3864938-2 ORF translation from 883-1326, direction F
VPLVILMIGMLAGSISHQVMHWGTFLLATTPIMLVAGKPYIQSAWASFKKHNNMDTLVAL
```

GTLVAYFYSLVALFAGLPVYFESAGFILFFVLLGAVFEEKMRKNTSQAVEKLLDLQAKTA
 EVLSDDSYVQVPLEQVKVRDLDDSSASR*

Description:

ATCS_SYNP7

Assembly ID: 3864956

Assembly Length: 1252bp

>[SEQ ID NO:73] 3864956 Strep Assembly -- Assembly id#3864956
 ACAAGAACAAATTGGAACAGGTACAGGCTGTTAAAAAATCGATTAAACACAGCTAGTGAAGA
 AGTGAAAAACCAAGTCTTGCTACCCATGGCTGATCACTTAGTGGCTGCTACTGAGGAAAT
 TTTAGCGGCTAATGCCCTCGATATGGCAGCGGCTAAGGGGAAAATCTCAGATGTGATGTT
 GGATCGTCTTTATTTGGATGCAGATCGTATAGAAGCGATGGCAAGAGGAATTCGTGAAGT
 GGTTCCTTACCAGATCCAATCGGTGAAGTTTGTAGAAACAAGTCAGCTTGAAAATGGTTT
 GGTATACAAAAAACGTGTAGCTATGGGGGTCATCGGTATTATCTATGAAAGCCGTCC
 AAATGTGACGTCTGATGCGGCTGCTTTGACTCTTAAGAGTGGAAATGCGGTTGTTCTTCG
 TAGTGGTAAGGATGCCTATCAAACAACCCATGCCATTGTACAGCCTTGAAGAAGGGCTT
 GGAGACGACTACTATTCATCCAAATGTGATTCAACTGGTGGAGGATACTAGCCGTGAAAG
 TAGTTATGCTATGATGAAGGCCAAGGGCTATCTAGACCTTCTCATTCCCTCGTGGAGGAGC
 TGGCTTGATTAATGCAGTAGTTGAGAATGCCATTGTGCCTGTTATCGAGACAGGAACTGG
 GATTGTCCATGTTTATGTCGATAAGGACGCAGATGACGACAAGGCACTGTCTATCATCAA
 CAATGCCAAAACAGTCGTCCTTCTGTCTGCAATGCCATGGAGGTTCTGCTGGTTCATGA
 AGACAAGGCAGCAAGCTTCCTTCCTCGCTTGGAGCAAGTGCTGGTTGCAGATCGAAAAGA
 AGCTGGGTTGGAACCAATTCAATTCCGCCTAGATAGCAAAGCAAGCCAGTTTGTTCAGG
 TCAAGCTGCTCAAGCACAAGACTTTGATACCGAGTTTTTAGACTATATTCTAGCTGTTAA
 GGTGTGTAGCAGTTTAGAAGAAGCGGTTGCGCATATTGAATCCACAGTACCCATCATTCCG
 GATGCTATTGTGACGGAAAATGCTGAAGCTGCAGCATACTTTACAGATCAAGTGGACTCT
 GCAGCGGTGTATGTTAATGCCTCAACTCGTTTCACAGATGGAGGACAATTTGGTCTTGGT
 TGTGAAATGGGGATTTCTACTCAGAAATTGCACGCGCGTGGTCCAATGGGCTTGAAAGAG
 TTGACCAGCTACAAGTATGTGGTTGCTGGTGATGGGCAGATAAGGGAGTAAG

ORF Predictions:

ORF #	Start	End	Direction	Length
1	1030	1251	F	74 aa

>[SEQ ID NO:171] 3864956-2 ORF translation from 1030-1251, direction F
 VTENAEAAAYFTDQVDSAAVYVNASTRFTDGGQFGLGCEMGISTQKLHARGPMLKELTS
 YKYVVAGDGQIRE*

Description:

gamma-glutamyl phosphate reductase (proA) homolog - Haemophilus
influenzae (str ain Rd KW20)

Assembly ID: 3864958

Assembly Length: 1785bp

```
>[SEQ ID NO:74] 3864958 Strep Assembly -- Assembly id#3864958
CTGCCCTAGCAGGAACGCAAGAAGGAAGTGGAGAATAGGCATTTTCAAAATTATAACCTA
CACTAGCCATCATATCTAATGTTGGAGTGCTAACTAGCTTATCCTTACTATTCAAGGATA
AGGCGTCTGCTCTCATTTGATCTACAACAATCAAAATAATATTTGGTTGTTTGTCTGAA
CCATAAAATCTCCTTTCTAATATGGCAAAAGAGGCACAAGAAGATATCTACCTTTACTGC
ACCCCTTTCTATATCAATCTCTCTATATAAAGCAATAACATTCTTGTTATGTTTTATAGA
ACAATGGACTAAAATATGACTAAATCGATTAGGAAATTCAAATCATTTTCTAGTACTGTT
TTAGTAAGTTACAGTGTACTATTCCAACCTCAATAAATTATAAACCTTTGTCTAATAACA
ATTTTAGTGGAGATAAGAAATCCTACACCTAACTCATCTTACACGTAATCTATTTCTATT
TTATCACAAAAACGCAAGTAAGACCATTAACCTCAATTCAGTTTTATCTGCCATTTTCAC
AAATGGGAAATAAGTCAAGACACTAATAATCAAACAAACAAGTATAAGATGATGGCAGC
CCAATCAAATGCTGTAGAGAAGAAACCATATAAAATTGGAGGCATTACCCAAGTAACATT
TTGTGTAACAGGTGAAACAAGACCCAGCTTGTTGCCAGTAAGCTACCGTTGCCATGAA
AACC GGGCTAAGTACAAATGGTATAAATAGCAAAGGATTCAAGACAAGTGGTAAACCATA
ATTCGATACCGGCTCACCAATATTAACAGAACTGGTGCTAGACCAAGTTTAGCAACTTT
TCGATAATGACTGTTTCTTGAAAAAATTAAAATAGCAAGTACTAATCCTAATCCTCCAAA
CCAGACAAACGCCCCAAAAGACCCACTTGTTCCATATATAAGGAATCGGTTTACCTTTTTG
GAAAGCATCCAGATTGCTAACATAGCAACTCCAAATAGCCCTTCCATGATGGGAGCCAA
TACATTTCTCCTCCATGGAGACCAAAAAACCAGAATAACTTATTCAAAAAGATCATCAGAAT
AACTGCAAAGAACTTTGAGACAAACCTAGTAATGGCGTTTGTAACACCTTGTAACCCCA
ATCAATCAATAAGTCATTGCTAAGTAAATGGAAAACATAAGTCAAGATGGCTACTATATA
CATCGCCATAAATCCTGGAATGATAGAAGTGAACGGCTTAGCAATCGCAGGGGGAAGTGA
ATCTGGTAACTTGATTACCCAGTTCTTTTTTATTACTTTACAGAAAATAATAGAGGCTAA
AAATCCAATCATCATGGCTGTAAAGTAGCCTCTGGCATTAAATATGGTTTCTGGAATCAC
ATTCCTCAATAGTTACCATCAGATTTTTTACCATCAAATGCTAGATTATCAATTCCATGTTA
AGATTTGATCTAATTTACATCTCCTACATTTGCCAAAGGGAACTCTTTGTAAGTGTAC
TTCCAATCGAAATGACAAACGAAGCAAGTGATACCAAACCAGCAGAACTGTATCAACCT
TGTAATCTTAGCGATATTCACTCCCAAGCAATAGATGAACAACAAGGAAACAATTGGTA
TACTTCCCTTGAATACCAAATTATTGATGTCAACAAGCCACTGAAAGGTTTTTCGTAATAC
TTCCTAGGTGAAATTGTTGTGGTAAATCCACTAGAAAAGCATTTAATAACAAAGCAATGG
AACCTGTCATAATAACAGGCATAGTCCCCACAAATGAATCACGTT
```

ORF Predictions:

ORF #	Start	End	Direction	Length
1	1427	1711	R	95 aa

Description:
unknown

[SEQ ID NO:75] 3865022 Strep Assembly -- Assembly id#3865022

ATCGAATTTTCATTTCTATTTTCCTATTCCATTTTTTATTCAAAAAATCAAAAAGCAAAGCTAG
AAAGCTGGTTCGCTGGTGGTTCAAAACACTGTTTTGAGATTGTCAATAGAACTGACAAACC
CTGTAATATACCTGCATATATACATACGACAAGGCGATACTACCCTAGTTTGAAGAGATT
TTCGAAGAGTATTCATTTTTTGTCTTTTACTTATTATACCATATTCACATAAAAAAACGAA
CATTCTTATCCTAAAAAATGCTCATTTTTTCTTAAATTATCAATCTAAATCTGGTTTATAG
AAGGAACGATTATCCATAGCGAAGATTTTATTGGTCATCTCTCCTTTATCCACCAAAGCC
AGAGCTGTTGACATCATCATCATGCTTGCATCCAGATTGTCAATCATATGGATAATCTCT
GCCTCCATAATACGTGGACGGACTGGAATTTCCATATTC AAGCAAGCCGTGGTGGACTTG
AGGATGACATGACGAAGCAAACGACTTCTTCCTTGGTATCATCGATGCCGAGTTCCATA
ACTGTCTTGGTAATTTTCGCTATCAATGAGAGCGATATGTCCAAGAAGATTACCTCGCACT
GTGTACTCTGTCTGGTCTGGCCCCGTCAACTCGATAACCTTAGCTAAGTCATGCAGCATA
ATCCCCGCATAGAGCAGGCTCTTATTGAGCTGAGGATAAACTTCGCTAATAGCGTCTGCC
AAACGTACCATGGTTCGCCGTATGATAAGCCAACCCCGTTTTCAAAGGCATGGTGGTTGGTC
TTGGCGGCTGGATAGGAGTAGAATTCTTATCATACTTGGTGTAGAGATTTTCGGACAATC
CGTTGCCAGACAGGATTTTCAATTTTGA AAATCATTTGCGACATGTAGTCACGAATTTCC
TTGACATCAACTGGTGACTTGACCTTGAAATCAGCTGGGTCATTGGGTTACCAGCTTGA
GGCAGGCGGAGAGTAATTTGATTGACTTGAGGGGTATTGTTATAAACTTCTCGGCGTCCT
TTCATGTGGACAACCTTACCTGCGGTAAAGGCCTCAATGTTATGAGGTTGGGCATCCCAG
AGCTTCCCATCAATCTCGCCACTATCATCTTGGAAGGTAAAGGCTAGGTAGTTTTTTCCCA
GCTCGAGTTTGCCTCAGGTCAGCTGATTTGATTAGGTAAAAGCCTTCAAATAACTCATCT
TTTTTTCATGTGACTAATCTTCATATTCTTCCTCATTTTCTTGAAAATGGAGTAGATCAAG
CGCAGGCTCACCTTCTGACAACTCAATGTGACGGAGCGTCCGCTCGATAGCTATGGTACG
ACGGTTTAATAAATTCGATCAATATTGCCAGAGGCATGTTGGAGATGTTTTTGTGCCTTGA
CCAGAA

ORF #	Start	End	Direction	Length
1	279	1271	R	331 aa

108

DGKLWDAQPHNIEAFTAGKVVHMKGRREVYNNTPOVNQITLRLPQAGEPNDPADFKVKSP
VDVKEIRDYMSQMIFKIENPVWQRIVRNLYTKYDKEFYSPAAKTNHHAFETGLAYHTAT
MVR LADAISEVYPQLNKSLLYAGIMLHDLAKVIELTGPDQTEYTVRGNLLGHIALIDSEI
TKTVMELGIDDTKEEVVLLRHVILKSTTACLNMEIPVRPRIMEAEIIHMIDNLDASMMM
STALALVDKGEMTNKIFAMDNRSFYKPDLD*

Description:

gi|710422 (U21636) cmp-binding-factor 1 [Staphylococcus aureus]

Assembly ID: 3865036

Assembly Length: 1167bp

>[SEQ ID NO:76] 3865036 Strep Assembly -- Assembly id#3865036
CTCAGATTACAGAGGACAATCAACTGGTTCATTTTCGTTTCCAGTTTCAAAAAGGCTTAG
AAAGGGAGTTTCATCTATCGTGTGGAAGAAAGTTAAGGCAGGTGTTCTCCTCTAC
GCAGTCACCATAGCAGCCATCTTTAGTCTTTTGTGCAATTTTATTTGAACCGACAAGTC
GCCCCTATCAAGACTATGCTTTGAATAAAGAAAAATTGGTTGCTTTTGCTATGGCTAAA
CGAACCAAAGATAAGGTTGAGCAAGAAAGTGGGGAACAGGTTTTAATCTAGGTCAGGTA
AGCTATCAAAACAAGAAACTGGCTTAGTGACGAGGGTTCGTACGGATAAGAGCCAATAT
GAGTTTCTGTTTCTTCAGTCAAAATCAAAGAAGAGAAAAAGAGATAAAAAGGAAGAGGTA
GCGACCGATTCAAGCGAAAAAGTGGAGAAGAAAAATCAGAAGAGAAGCCTGAAAAGAAA
GAGAATTCCTAGTCAATTCAACTATAATGCGTTGAATCCAGAATAGTCCACTGTAGTTTC
TAGAAAATTGCTGGAAATGGATGTTAAGCTCCAATTCATTTGTTTATATCTTATTTTCAGT
CCACTATACTTTGTGCTAAATTAAAGATATGAAACATGATTTTAACCACAAAGCAGAAAC
TTTCGATTTCCCTAAAAATATCTTCCTCGAACTTGGTATGTCAAGCAGCCGAGAAACA
GATTGATCTTCTATCAGACAAAGAAATTTTAGATTTTCGGTGGTGGCACGGGTCTATTAGC
CTTGCCCCCTAACCCCTAGCCAAGCAGGCTAAGTCAGTCACTCTTGTAGACATTTCTGAGA
AAATGTTGGAGCAAGCTCGTTTGAAAGTGGAGCAGCAAGCAATCAAGAATATCCAGTTT
TGGAGCAAGATTTACCGAAAAATCCCTTGGAGAAAGAGTTTGATTGCCTTGCTGTTAGTC
GGGTTCTTCATCATATGCCTGATTTGGATGCGGCTCTCTCACTGTTTCATCAACATTTGA
AGGAAGATGGGAACTCATCATTGCTGATTTTACCAAGACAGAAGCTAATCATCATGGAT
TTGATTTAGCTGAACTGGAAAACAAGCTAATTGAGCATGGGTTTTTCATCTGTGCATAGT
CAGATNCTCTATAGCGCTGAAGANCTG

ORF Predictions:

ORF #	Start	End	Direction	Length
1	79	492	F	138 aa

>[SEQ ID NO:174] 3865036-1 ORF translation from 79-492, direction F
VWKKKKVKAGVLLYAVTIAAIFSLLLQFYLNQVAHYQDYALNKEKLVAFAAMAKRTKDKV
EQESGEQVFNLGQVSYQNKKTGLVTRVRTDKSQYEFLLPSVKIKEEKRDKEEVATDSSE
KVEKKKSEEKPEKKENS*

Description:

unknown

Assembly ID: 3865054

Assembly Length: 916bp

```
>[SEQ ID NO:77] 3865054 Strep Assembly -- Assembly id#3865054
TCTCCCAACATATAATTTCCGTTTTTCCAATCCCCCAGCTGTCATACAGTCTGTGATAAGA
GCGATGTTTTCTGTTCCTTTTTGTTTTGATAAGAATTTTCGCAAGCCTTTGGATCTACGTGG
TGACCATCACAGATCAACTCTGCATAGGTATGTGGCAATTGGTACATGGCTCCAACCATA
CCCAATTCACGGTGAGTCAACCCACGCATTCCATTGTAGGCATGCACCCAAACACTCGCT
CCAGCATCGACTGCTTTTTTTGGCTTCATCAAAAGTCGCGTTTGAATGTCCAAGAGCAACC
GTCACACCTTCGCCCCTAACTGTACGAACAAAGTCTTCCACCCCATCACGTTCTGGTGCA
ATCGAATTTTATTAAGCAAGCCATTTGCCGCTTTTTTGCCAAGAATGAAACTCCTCAACAC
CCGGGTCTCTCATATAAGTTGGATTTTGTGCCCCCTTAAAAGTTTCTGTGAAATATGGAC
CTTCATAATAAATCCCACGAATCTTAGCACCTGTTGCTTCTTTATAATGGTTTCCAAGAT
TTTCAGTGAAGCAATTGCTCATAAGTGGCTGTTAAAGTTGTGGGTAAGAACTGG
TAACACCGGTACTAAGAAGTCCTTCACTCATAGTATGCAATGTACCTTCAATGTTGTTGT
CCATCACATCTACACCTGCATATCCATGAATATGAGTATCCACAAGACCTGGGGCAATGC
TATAACCTGTATAGTCAATCACCTCAGCCCCTTCAGGAATCTGCTCTACATGTTTCCCAA
ACTTGCCGTCCACAAGTTCCAAGTAACCACCTCGACAAATCCGTGTGGGTTAGAAAACTG
ATCCGCTTTAATATAGTTAGGCATAATGTTAACCTCCTTAAAAGATTGATTCTACAATTT
ATTATGTCAATTCGAT
```

ORF Predictions:

ORF #	Start	End	Direction	Length
1	302	793	R	164 aa

```
>[SEQ ID NO:175] 3865054-1 ORF translation from 302-793, direction R
VDGKFGKHVEQIPEGAVIDYTGYSIAPGLVDTHIHGYAGVDVMDNNIEGTLHTMSEGLL
STGVTSFLPTTLTATYEQLLAVTENLGNHYKEATGAKIRGIYYEGPYFTETFKGAQNPTY
MRDPGVVEEFHSWQKAANGLLNKIRLHQNVMGWKTFLVQLRAKV*
```

Description:

N-acetylglucosamine-6-phosphate deacetylase (nagA) homolog -
Haemophilus influe nzae (strain Rd KW20)

Assembly ID: 3865102

Assembly Length: 786bp

```
>[SEQ ID NO:78] 3865102 Strep Assembly -- Assembly id#3865102
CTGGATTAAAACGAGGCAGTTTCAGACTAATATCCAAGTCGTAAGAAATGCCTGAAATAA
```


GCTTTTCTAAATTGTCCAAAGCTTGCGGGAAAACGCTCTTGGAATAGTTTCTCTAAAGAA
 CTTGCTGATATAAAGACATCTTGTCTCGAACGCAAGGGAACCTTCTCTGAGCGGTAGATTT
 TCTTTAATCGCTGTTAAAACCTTGAAGAACTTCTCTATCCCTGCTTTCAAAGCGTTGACC
 CGATAAAGAGGTAAGATAGGATGATGAAATTTCGCTTGCTAGTGTTTCTGGATAAACCCCT
 ATATAGTAATCACAGCCTAGTTCTAACGACTCAACTCTATCAAATAAGGCACAATGACC
 GCGATATCCTCCAGGTACTGGGACAGGACTGACCAAGTTTTCTCCCCCTGCATCTTGGCT
 GTCGAAAGCTTCATCAACTGCTGATAGCCCACACTAGATAGAGCTAAAAAGCGCAAATTC
 ACTTCCTGATCATCTACAAACACTGTCATTTCAAGCCCTAGCAAAGGATGAATGCCGTAT
 TTTTTTGTAATCTCTAGAAAGTCGAAAGCGCCATAAAGATTGTCAATATCCATCATAGCC
 AAATGAGTGTAGCCGATTCTTTAGCTGCTCTCACATACTTTTCGATCGAAATGACGCTT
 TCCATAAAACTATAGACTGTTTTTGTATCTAGTTGTGCGATCAATTTACACTTCTCCTCT
 ATCCTTCTCACTATATTATACCATTTTCACCTATAAATGGCTTCTCTTGAGAAAAATTC
 GATCAG

ORF Predictions:

ORF #	Start	End	Direction	Length
1	27	731	R	235 aa

>[SEQ ID NO:176] 3865102-1 ORF translation from 27-231, direction R
 VRRIEEKCKLIAQLDTKTVYSFMESVISIEKYVRAAKEYGYTHLAMMDIDNLYGAFDFLE
 ITKKYGIHPLLGLEMTVFVDDQEVNLRFLALSSVGYQQLMKLSTAKMQGEKTWSVLSQYL
 EDIAVIVPYFDRVESLELGCDDYYIGVYPETLASEFHHPIPLYRVNAFESRDREVLQVLT
 AIKENLPLREVPLRSRQDVFISASSLEKLFQERFPASFGQFRKAYFRHFLRLGY*

Description:

unknown

Assembly ID: 3865156

Assembly Length: 1213bp

>[SEQ ID NO:79] 3865156 Strep Assembly -- Assembly id#3865156
 CACTTTCAGCTTCTTCTTTTTGAACGGTTATAAACACGAATCAGATTCCTTATTTCTT
 GCGATTTATGTGATTCCTTATTTTCCAATCTAAAGTATAGTGAAATGAAATAAACATGC
 GCAAATCGATTAAGGAATTTAATCTAATTTCTAACAATGTCTTAGAAATCAAAGTGACT
 ATTTTAACTTCAATGCACTAAACATCTAATACTCAATAAAAAATCAAAGAGCAAAC TAGGA
 AACTAGCCGCGAGGTGGCTCAAAACACTGTTTTGAGGTTGTAGATGAAACTGACGAAGTCA
 GTAACCATACATACGGCAAGGCGACGCTGACGTGGTTTGAAGAGATTTTCGAAGAGTAGC
 AAAATGGAAAAAGGAGTGAGTGAAGCACATCGCCTCCCCACTCCTTTTTCTGTTTTTAGG
 CTGTTTTTTTCAACCTTCAAGATTTTACATCATAGCTACCAACAGGCGTTTCAATGGTTG
 CTGTATCACCTGTTTTCTTGCCAATCAAGGCCTGCCCAATTGGGCTTTCATTTGAAACCT
 TACCTGCAAAGGCATCCGCACCAGCTGAACCTACGATAATATAAACTTCTTCTTCGTCCT
 CACCAATTTCTTGGATGGTGACTGTTTTACCAATCGCTACTTCGTCCTGGGCAACTGCGT

CGCTATTGACGATTTTCAGCATAGCGGATTTTTGTTTCTAAGCTAGAGATTTGTCCTTCGA
 CAAAGGCTTGTTTCATCCTTAGCTGCTTCGTACTCACTGTTTTCTGAAAGGTCACCGTATG
 AACGGGCAATCTTAATGCGTTCTACCACTTCTGGTCGACGAAACCAATTTCAATTCTTCT
 AATTCTTTTTCAAGTTTTTCTTTTCTCAAGGGTCATAGGATATGTTTTTCTGCCATT
 TTTCTCAACTTTCTTCTGATAATATTTTCTAAAGAAAATTATGTGAAGTATCACATAATT
 TTAGTTTGTTTAGTTTAATTTGCTGTTGACATGTTTCAGCGACATTGCGGGTCGTGGTCTTC
 TTGATTGTTAGCATAGTAAACCTTGCCTTCTGTGACATCTGCTACAAAGTAAAAGTTATC
 GCTCTTAGTTTGATTGATGCTTGACTCAATCCGCATCCAAGACTTGGACTATCGACTGGA
 CCAGGCATGAGACCTACATTTTTATAAACATTATAAGGTGAATCAATGTTGGTATCAATC
 GCAACATCCTCAG

ORF Predictions:

ORF #	Start	End	Direction	Length
1	416	808	R	131 aa

>[SEQ ID NO:177] 3865156-1 ORF translation from 416-808, direction R
 VVERIKIARSYGDLSSENSEYEAAKDEQAFVEGQISSLETKIRYAEIVNSDAVAQDEVAIG
 KTVTIQEIGEDEEEVYIIVGSAGADAFAGKVS NESPIGQALIGKKTGDTATIETPVGSYD
 VKILKVEKTA*

Description:

TRANSCRIPTION ELONGATION FACTOR GRE A (TRANSCRIPT CLEAVAGE FACTOR
 GRE A). - ESCHE RICHIA COLI.

Assembly ID: 3865160

Assembly Length: 1173bp

>[SEQ ID NO:80] 3865160 Strep Assembly -- Assembly id#3865160
 TGGCGCTGAGTTGGGAATTCCTATCGTTAATAAGCGTGTATCGGTGACACCTATTTCTCT
 GATTGGGGCAGCGACAGATGCGACGGACTACTGGTTCTGGCAAAAGCGCTTGATAAGGCT
 GCGAAAGAGATTGGTGTGGACTTTATTGGTGGTCTTTCTGCCTTAGAACAAAAAGGTTAT
 CAAAAGGGAGATGAGATTCTCATCAATTCCATTCCTCGCGCTTTGACTGAGACGGATAAG
 GTCTGCTCGTCAGTCAATATCGGCTCAACCAAGTCTGGTATTAATATGACGGCTGTGGCA
 GATATGGGACGAATTTATCAAGGAAACGGCAAATCTTTCAGATATGGGAGCGGCCAAGTT
 GGTGTATTTCGCTAATGCTGTTGAGGACAATCCATTTATGGCGGGTGCCTTTCATGGTGT
 TGGGGAAGCAGATGTTATCATCAATGTGCGGAGTTTCTGGTCCTGGTGTGGTGAAACGTGC
 TTTGGAAAAAGTTTCGTGGACAGAGCTTTGATGTTAGTAACCCGAAAACAGTTAAGAAAA
 CTGCCTTTTAAAATCACTCCGTATCCGGTCCAATTGGTTTGGTCAAATGCCAGTGAGAG
 ACTGGGTGTGGAGTTTGGTATTGTGGACTTGAGTTTGGCACCAACCCCTGCGGTTGGAGA
 CTCTGTGGCACGTGTCCTTGAGGAAATGGGGCTAGAAACAGTTGGCACGCATGGAACGAC
 AGCTGCCTTGGCCCTCTTGAACGACCAAGTTAAAAAGGGTGGAGTGATGGCCTGTAACCA
 GGTCGGTGGTCTATCTGGTGCCTTTATCCCTGTTTCTGAGGATGAAGGAATGATTGCTGC

AGTGCAAAATGGCTCTCTTAATTTAGAAAACTAGAAGCTATGACGGCTATCTGTTCTTG
 TTGGATTGGATATGATTGCCATCCCAGAAGATACGCCTGCTGAAACTATTGCGGCTATGA
 TTGCGGATGAAGCAGCAATCGGTGTTATCAACATGAAAACAACAGCTGTTTCGTATCATTC
 CCAAAGGAAGAGAAGGCGATATGATTGAGTTTGGTGGTCTATTAGGAACTGCACCCGTTA
 TGAAGGTTAATGGGGCTTCGTCTGTCTGACTTCATCTCTCGCGGTGGACAAATCCCAGCAC
 CAATTCATAGTTTTTAAAAATTAAGAAAATAGGA

ORF Predictions:

ORF #	Start	End	Direction	Length
1	136	375	F	80 aa

>[SEQ ID NO:178] 3865160-1 ORF translation from 136-375, direction F
 VDFIGGLSALEQKGYQKGDEILINSIPRALTETDKVCSSVNIGSTKSGINMTAVADMGRI
 YQGNKSFYRYSQVGCIR*

Description:

unknown

Assembly ID: 3865172

Assembly Length: 1209bp

>[SEQ ID NO:81] 3865172 Strep Assembly -- Assembly id#3865172
 TCGGAATCTGAGCTAGTGTAGCTTCCTTAATCTTATCTGATAAGATAGCTGTCATATCAG
 ACTCAATCATTTCTGGAGCAATCAACATTGACTCGTATATTCGACTAGCGACCTCGCG
 TGCCACAGACTTGGTAAAGCCAATCAAGCCAGCCTTAGAAGCAGCATAGTTAGCTTGACC
 AATATTCCTCATCAAACCAACAACACTAGACATATTAATGATAGCACCTTCTCTGGCTTT
 CATCATCGGTTTCAAGACTGATTGTGTTCATATTAAAGGCACCAGTCAGATTGACCTTGAG
 CACTTTTTTCAAATCTGCTTCTGTTCATCTTGAGCATAAGAGTATCTTGGGTAATCCCTGC
 ATTGTTGACCAAAACATCTACTGAACCCAGTTCTGCAATAGCTTGATCAATCATACGCTT
 AGCGTCTGCAAATCTGATACATCTCCTGAAATGGGAACCACCTTGATACCATAGTTTGA
 AAATCAGCGAGCAATTCTTCTGAGATTGCCCCACGACTGTTTAAGACAATGTTGGCTCC
 TGCTTGAGCAAACCTTGTGGGCGATGGCAAGACCAATTCCACGACTCGAACCTGTAATAAA
 GATATTTTTTATGTTCTAGTTTCATTTTTTTCCTTTCAAACCTTCTACTTATTTTAGTCTA
 TTTTTCTAAAGTGCTACTAACTCGCTTGATCTTCCACATGAGCTAAGTGAGCAGTTTG
 ATCAATTTTTTTAACAAAACCTGACAAGACTTTCCCCGGTCCAATCTCGAATAAAGTTGC
 TTATGCCTGCTTCTTGCATGACCCCAATACTTTCATAGAAACGAACGGGTTCCCTTGACCT
 GACGCGTCAAGAGCTGAGCAATGTCTCTTTTTGTCATCACAGCAGCTTCTGTATTGCCGA
 CTAGGGGACAAGTAAAATCTGAAAACTTACCTGAGCTAGAGTTTCAGCTAGTTTCTGGC
 TAGCAGGCTCAAGGAGAGCGGTGTGAAAGGGACCTGACACCTTAAGAGGAATCAAGCGTT
 TGGCACCTGCTTCTTGCAAAAGTTCAACCGCTCGATCAACTGCAACCACTTCTCCAGCAA
 TGACGATTTGTGCAGGTGTGTTATAGTTGGCTGGAGTAACCACTCCAAGTTCCAGAAGCT
 TTTTGACAGGCTTCTTCAATGACCTCTACTGGCGTATTGAGAACTGCTACCATCTTGCCA

AGTTCAGCA

ORF Predictions:

ORF #	Start	End	Direction	Length
1	731	1123	R	131 aa

>[SEQ ID NO:179] 3865172-2 ORF translation from 731-1123, direction R
 VVTPANYNTPAQIVIAGEVVAVDRAVELLQEAGAKRLIPLKVSGPFHTALLEPASQKLAE
 TLAQVSFSDFTCPLVGNTAAVMQKEDIAQLLTRQVKPVRFYESIGVMQEAGISNFIRD
 WTGESLVRFC*

Description:

malonyl coenzyme A-acyl carrier protein transacylase (fabD) homolog -
 Haemophilus influenzae (strain Rd KW20)

Assembly ID: 3865228

Assembly Length: 813bp

>[SEQ ID NO:82] 3865228 Strep Assembly -- Assembly id#3865228
 ATGACACGTCTGTTCTCTCAAGCAGAAATGGCAGAGTAACAAGCTCGATATTGAGGTAGC
 CGATAAAGAATTGGCTGAATTTGAAGCTCAGATTAAACAGGAAGTGGAAGCTCCAATTG
 TAGTGAGTCCTCAGGTTGAAGAAGAGCCTCAGCTCATCCAGTTGGCCCAATGTATGAAGA
 ACCAGAAGTAAATCCAGTGCATCCGACAGGTCCAACACCAGCTACAGAACTGTTGATTC
 AATACCGGGATTTGAAGCACCGCAAGAATCTGTTACAATTTTATAAGAAATATTCTGAGA
 ACAATATCTTATCCTTATATTTCCAGCGAGCAGGAAATGGTGTGAGTCCTGCATTCCCTA
 TCGATAAGATTATCCTCTCAAACTATCAAGTCTGAATCTAGTAAGATTTGACGTTCCCCA
 CGTTACGGGATAAGAGAGAGAAAGACTAAATCTTTTCCGAATAAAGGTGGTACCACGAT
 TTTTCGTCCTTTTGGGAAGTCGTGGTTTTAAATTTGTTATTATTTATAAAGGAGATACCAT
 GAAACTCAAAGACACCCTTAATCTTGGGAAAAGTGAATTCCCAATGCGTGCAGGCCTTCC
 TACCAAAGAGCCAGTTTGGCAAAAGGAATGGGAAGATGCAAACTTTATCAACGTCGTCA
 AGAATTGAACCAAGGAAAACCTCATTTACCTTGCATGATGGCCCTCCATACGCTAACGG
 AAATATCCACGTTGGACATGCTATGAACAAGATTTCAAAGATATCATTGTTTCGTTCTAA
 GTCTATGTTCAGGATTTTACGCGCCATTTATTC

ORF Predictions:

ORF #	Start	End	Direction	Length
1	197	286	F	30 aa

>[SEQ ID NO:180] 3865228-1 ORF translation from 197-286, direction F
 VHPTGPTPATETVDSIPGFEAPQESVTIL*

Description:

unknown

Assembly ID: 3865230

Assembly Length: 953bp

```
>[SEQ ID NO:83] 3865230 Strep Assembly -- Assembly id#3865230
ATCGAATTATTTTGAACAAGGTGGATCAGCTATTTTGGCCTTGATTAGTATTTTACTCT
TTAAATACACTTGAAGGTCGATTCTAATCTCGCTAATCCTTTTAAATCCAGAATAAGGGA
AATATGTTATACTTGTTTTTAAGAAAAAGTTTCATTGAATTGGTTTTGAGGAGTTAGAA
ATGAAAGTATTAGTGACAGGTTTTGAGCCCTTTTGAGGCCATTAAAGGTTTACCAGCTGA
AATCCATGGTGCTGAGGTCCGTTGGCTAGAGGTGCCGACAGTTTTTCACAAATCTGCTCA
AGTATTGGAAGAAGAGATGAATCGTTATCAACCTGACTTTGTCCTTTGTATTGGGCAAGC
TGGTGGAAGAACTAGTTTGACACCTGAACGAGTGGCCATTAATCAAGACGATGCACGTAC
TTCTGATAACGAAGATAATCAACCGATTGACCGTCCCATTGCGCCAGATGGTGCTTCGGC
CTACTTTAGTAGTTTGCCGATTAAAGCGATGGTTCAAGCTATAAAAAAGAAGGATTACCG
GCCTCTGTTTCCAATACGGCAGGGACTTTTGTCTGCAGCCATTTGATGTATCAGGCTCTC
TATTTGGTAGAAAAGAAATTCCCATATGTTAAGGCAGGTTTTATGCATATTCCTTATATG
ATGGAACAGGTGGTGAACAGACCGACTACTCCAACCTATGAGTTTAGTGGATATTCGGCGA
GGGATAGAAGCAGCAATCGGCGCTATGATAGAACATGGAGATCAGGAACCTCAAGTTGGTA
GGCGGAGAAATTCATTGATAGAAAAAGCTTGAGGGGAAAACCTTCAAGCTTTTGGACGT
TTTCGAGCCAATACTGCTCGGTAAAACATAATTTTAGTGCATTGGATATAAGGTAGGAGT
GAAAAACTAGCAATGCCAAAGGTAATCCAATTGAGGAAGTACCAAGGAAGAAG
```

ORF Predictions:

ORF #	Start	End	Direction	Length
1	272	586	F	105 aa

```
>[SEQ ID NO:181] 3865230-1 ORF translation from 272-586, direction F
VPTVFHKSQVLEEEMNRYQPDFVLCIGQAGGRTSLTPERVAINQDDARTSDNEDNQPID
RPIRPDGASAYFSSLPIKAMVQAIKKKDYRPLFPPIRQGLLSAAI*
```

Description:

PYRROLIDONE-CARBOXYLATE PEPTIDASE (EC 3.4.19.3) (5-OXOPROLYL-PEPTIDASE). - STR EPTOCOCCUS PYOGENES.

Assembly ID: 3865378

Assembly Length: 1060bp

```
>[SEQ ID NO:84] 3865378 Strep Assembly -- Assembly id#3865378
CTACTTGAAACAGAACTGAAATTATACCCACTACCTCCCTGATTATCTTCAATGCTTACG
TCTAAATAAACTTCCCCACTATTATTTAGCTTAGCAACAACCTGTTATAGTAAATAACAT
```

AAAATTCACATAAATAGATTAGGGAAATCAAAGCAACTTCTAGGAATGTTTTAGCAGTCA
 CAGTGTACTTTTCCAGCATCAAGCCACTATAACTCTGCACATAAAAATGGAGAAGATGGC
 CATCCTCTTCTCCAAATATTAACCTTCTTTACAAACCAACTATAGTTGACAAAGAACCTAA
 AATCAATTGATAACACGAGGTCAGGTCGGTCAACTCTTTCAACTGAAGCCCTGTCAACTC
 TTCCCATTTATCAATCTTGTATTGGAGAGAATTGCGGTGCAGATAGAGTTGCTGGGCTGT
 TTAAGTGAGAACAGCACTATTTTCCCAAAGAGAGAGAATGATTTCCCTGAATCTGATCTTG
 ATCCAAAATCATCTGGTGTAGACATTCTTGATTGGCTTCAAGTCCACGAGTCTTTCTCC
 CAGACTCCAAAGATAGAGCTGAGAAAAAGTATGAACACCTTGGTGACCCTGACGCCACCA
 TGTCTTGAACAAATCCCGCTCAGCTTTGATTAAGTCTGATAGGGCTTGATGTCCCGTCTG
 AGACCAAACCTGACCCAACATGATAGAAAGACGAAGTCCAAAGTCATACTCAACCGCTTC
 AATCGTATCACTTAAATATCTCTTACAGAAGTGTATTTGTCTTGTTGAAGCACGAAAAC
 ATAATCCTGAGATCCGACCTGTAGCACTGTCTGACAATTCGAAAAAGAGTCCGCATCAT
 ATCTAGCCAAGAAGCCAGATTTTCCTGCTGAAAATAAGAAAGATGGCAATAAACCAACTG
 AATCTTTTAAAAACTTGCAGGTGCCTGTCCCTTGCTTCAACCAGATAGGAATACCAAGG
 GTTTAGCGAACGAACCTGCTCCTGCTGGGTCAAAGGGCAACCAACTGCTTTTCACGCTC
 GCTGAGCCCAGCTTCCTCCAGCAAAATCCACTGCTGAGAG

ORF Predictions:

ORF #	Start	End	Direction	Length
1	421	807	R	129 aa

>[SEQ ID NO:182] 3865378-1 ORF translation from 421-807, direction R
 VLQVGSQDYVFLVQLQDKYTSVRDILSDTIEAVEYDFGLRLSIMLGQVWSQTGHQALSDLI
 KAERDLFKTWWRQGHQGVHTFSQLYLWSLGERLVDLKPIKECLHQMILDQDQIQEIILSL
 WENSAVLT*

Description:

unknown

Assembly ID: 3865470

Assembly Length: 895bp

>[SEQ ID NO:85] 3865470 Strep Assembly -- Assembly id#3865470
 ATTTTAGACTTTGATGACAATCCTCAGGCGGTATCATGCCCAATCACGAGGGGCTGGAA
 TTGCAGTTGCCAAAGAAGTGTGTTTATGCATTTTGTAGGTGAGGAGATCTGACCGCTATGC
 AAGGGAAGTAGGGGCGGATTGTGTCGGCGAATTCGTTTCTGCTACCAAGACCTATCCAGT
 CTCTTTCATCAACTACAAGGGTGAGGAGGTCTGTCTGGATCAGGCTCCTGCTGGCTCCGC
 TCCAGCAGCCCAGTTTATGGATGGGTTGATTGGCTATGGTGTGGAGCAGCTTATCTCTAC
 TGGGACCTGTGGTGTCTAGCTGATATAGAGGAAAATGCCTTTCTAGTCCCTGTTCGCGC
 TTTGCGAGATGAGGGAGCCAGTTACCACTATGTGGCACCTTGTCGTTATATGGAAATGCA
 GCCAGAGGCTATTGCTGCTATTGAGGAAGTTTGGGAAGACAGAGGGATTCTTATGAAGA
 AGTCATGACCTGGACGACAGACGGTTTTTACCAGAGAAACGGCTGAAAAGGTGGCTTATCG

TAAGGAAGAAGGCTGTGCTGTTGTGGAGATGGAGTGTCTGCTCTTGCGGCAGTAGCTCA
 ATTGCGTGGGGTTCTCTGGGGTGAATTGTTGTTCACAGCAAATTCTCTAGCGGACTTGGA
 CCAGTACAACAGTCGTGACTGGGGCTCGGAACCTTTTAATAAGGCGCTAAAACTGAGTTT
 AGCAAGTGTCCACCACCTTTAGTTGTACTGGCAAAGGATTTGTTTTATCATAAAATGTCT
 AGCTCATACTTTTCAAAAATATGTTTAAACGAAGTCACCTTCCTCTTGTCTTAAGCATGT
 TTGAAGTTGGGAAAAATCTTTAAATCAGAAAAACGTATCATATCAGGTTGATGA

ORF Predictions:

ORF #	Start	End	Direction	Length
1	98	742	F	215 aa

>[SEQ ID NO:183] 3865470-1 ORF translation from 98-742, direction F
 VRRSDRYAREVGADCVGEFVSATKTPVPSFINYKGEEVCLDQAPAGSAPAAQFMDGLIGY
 GVEQLISTGTGCVLADIEENAFVLPVRALRDEGASYHYVAPCRYMEMQPEAIAAIEEVLE
 DRGIPYEEVMTWTTDGFYRETAEKVAYRKEEGCAVVEMECSALAAVAQLRGVLWGELLFT
 ANSLADLDQYNSRDWGSEPFNKALKLSLASVHHL*

Description:

unknown

Assembly ID: 3865632

Assembly Length: 645bp

>[SEQ ID NO:86] 3865632 Strep Assembly -- Assembly id#3865632
 AGGGCTGTCAAGCTTGGTTAGAACGTTTAGAAAAGGAGAGTTAAGGTGGAAAATCTTACG
 AATTTTTACGAAAAGTATCGTGTCTATCTGACTCGTCCACGTTTAGAGCTTTTGGCAGTA
 GTTACCATTGTTTTANGNGCTGTACTCGTCTTTTTTCTAAATATTCCAGGAAAAGGTGTC
 TTAAACTCGATAATGGAACGATTGTTTATGATGGCAGTCTTGTCCGTGGTAAAATGAAT
 GGCCAAGGTACCATTACCTTCCAAAATGGAGACCAATATACAGGTGGCTTCAACAATGGA
 GCCTTCAACGGAAAAGGTACCTTTCAATCTAAAGAAGGCTGGACCTACGAAGGTGATTTT
 GTAAATGGTCAGGCTGAAGGAAAAGGGAACTAACAACAGAACAAGAAGTCGTTTATGAA
 GGAAGTTTAAACAAGGCGTTTTTCAACAAAATAAAGCCTCCTTATCAAAGGAGGTATT
 ATTAGAATTACAAGGTAAGCGTTTACCTGTAAATCCCTTTCTTTCCAAATCCCTCTTCCA
 AGCAAGTTTGTGAAATAAAAAATATTTGAAATAAATTTCAAACTTCAAAGATAAAACC
 TGATAAGAAAAGAAAATGAGAAAAGTTTCGCAAGAGTTTAAAAAT

ORF Predictions:

ORF #	Start	End	Direction	Length
1	46	456	F	137 aa

>[SEQ ID NO:184] 3865632-1 ORF translation from 46-456, direction F
 117

VENLTNFYEKYRVYLTRPRLELLAVVTIVLXAVLVFFLNIPGKGVCLKLDNGTIVYDGS
 LGKMNQGGTITFQNGDQYTGGFNNGAFNGKGTFFQSKEGWTYEGDFVNGQAEKGKLTTEQ
 EVVYEGTFKQGVFQQK*

Description:

unknown

Assembly ID: 3865710

Assembly Length: 572bp

>[SEQ ID NO:87] 3865710 Strep Assembly -- Assembly id#3865710
 GAGATCTGTCTTGACACCAAAAGTGTGGAGTACGCCAGCTAATTCAACGGCGATATAACC
 AGCGCCTAGAATCGCAATTGACTCTGGAAGTTCTTCCCAGGCAAATACATCATCAGAAGA
 GCCACCTAGCTCAGCACCAGGAATATTAGGAATACTTGGATGGGCACCTGTAGCAATCAC
 GATATGTCTAGCACGAATCAGTTCACCATTTACGCTTACAGTATGAGAATCTACAAATTC
 AGCATGACCTTCAATCAAGTCTACACCGTTGCGTTTAAAACTACCATCATAGAGAAGAAC
 GAGCGCGATCAATGTAGGCTTCACGATTGCGACGTAGGGTTGCAAAGTTAAAGTTAAGAT
 CAGTAGTCTCAAAGCCGTAGTCTCCTCCAAATTGATGGAAAGTCTCAGCGATTGCGCCCC
 CGCTACCACATGATTCTTTTAGGAACACAACCGACGTTGACACAGGTTCCACCTAATTTTC
 TTTTCCTCAATAACGGCTGCTTTGGCTCCATGTTCCCAGCACGGTTCATGGTAGCGATCC
 TCCGCTACCTCCACGATAGCAATGATATCATA

ORF Predictions:

ORF #	Start	End	Direction	Length
1	287	448	R	54 aa

>[SEQ ID NO:185] 3865710-1 ORF translation from 287-448, direction R
 VFLKESCGSGAQIAETFHQFGGDYGFETDLNFNFATLRRNREAYIDRARSSL*

Description:

glutathione reductase (NADPH) (EC 1.6.4.2) - Streptococcus
 thermophilus

Provided in Table 2 is information on the direction of the ORF (forward or reverse) for each polynucleotide in Table 1. Also listed for each ORF is its start and stop codon positions (refer to the columns containing nucleotide code labeled "Start" and "Stop"). The triplet codon sequence for each start and stop codon is also shown. These codons may be shown in the sense orientation or antisense orientation, such as GTG and CAC, respectively, for start codons. The "Length" column discloses the length of each polynucleotide assembly. The direction of translation on the polynucleotide depicted is denoted by and "Forward" for forward or and "Reverse" for reverse (or being on the opposite strand from the one depicted). As indicated above, the "Assembly ID" number is a unique identifier assigned to each ORF of Table 1 and allows a correlation between the data in Tables 1 and 2.

TABLE 2

Assembly ID	Start	Stop	Start	Stop	Length	Direction
3049156	~CAC	TCA~	236	385	50	Reverse
3049862	GTG	TGA	383	526	48	Forward
3112810	~CAC	TTA~	601	804	68	Reverse
3112866	~CAC	TTA~	220	513	98	Reverse
3113664	GTG	TAA	165	392	76	Forward

Assembly ID	Start	Stop	Start	Stop	Length	Direction
3113716	~CAC	TTA~	94	291	66	Reverse
3174176	GTG	TAA	139	543	135	Forward
3174186	GTG	TAG	83	283	67	Forward
3174374	GTG	TGA	154	294	47	Forward
3174972	~CAC	TTA~	169	678	170	Reverse
3175138	~CAC	TCA~	79	945	289	Reverse
3175860	GTG	TAA	51	251	67	Forward
3175918	GTG	TGA	212	535	108	Forward
3811220	~CAC	CTA~	316	873	186	Reverse
3811436	~CAC	TTA~	1164	1511	116	Reverse
3811984	GTG	TGA	134	454	107	Forward
3857228	~CAC	TCA~	1141	1356	72	Reverse
3857842	GTG	TAA	45	341	99	Forward
3857996	GTG	TAA	58	456	133	Forward
3858236	~CAC	CTA~	1	261	87	Reverse
3858264	~CAC	TCA~	439	1365	309	Reverse
3858610	~CAC	TTA~	374	949	192	Reverse
3858716	~CAC	CTA~	238	402	55	Reverse
3859124	~CAC	CTA~	73	453	127	Reverse
3859244	~CAC	TTA~	310	462	51	Reverse
3859250	~CAC	CTA~	244	402	53	Reverse
3859588	~CAC	TTA~	102	443	114	Reverse
3859774	~CAC	CTA~	9	131	41	Reverse
3860140	GTG	TAA	302	511	70	Forward
3860140	GTG	TAA	605	856	84	Forward
3860206	~CAC	TTA~	898	1056	53	Reverse
3860270	GTG	TAG	346	966	207	Forward
3860438	GTG	TAG	1	276	92	Forward
3860438	GTG	TGA	460	1128	223	Forward
3860544	GTG	TAA	222	689	156	Forward
3860558	~CAC	TTA~	717	1376	220	Reverse
3860568	GTG	TAA	1040	1291	84	Forward
3860582	GTG	TGA	356	1027	224	Forward
3860724	GTG	TGA	139	498	120	Forward

Assembly ID	Start	Stop	Start	Stop	Length	Direction
3860724	GTG	TGA	686	1024	113	Forward
3860858	GTG	TAG	610	807	66	Forward
3860890	GTG	TAG	397	486	30	Forward
3860952	~CAC	TTA~	449	715	89	Reverse
3860962	~CAC	TTA~	152	646	165	Reverse
3861268	~CAC	TTA~	457	645	63	Reverse
3861270	~CAC	TTA~	627	824	66	Reverse
3861288	~CAC	CTA~	357	572	72	Reverse
3861306	GTG	TAA	717	1208	164	Forward
3861306	GTG	TAA	1201	1410	70	Forward
3861334	GTG	TAA	76	975	300	Forward
3864148	GTG	TAG	212	940	243	Forward
3864148	GTG	TAA	1202	1753	184	Forward
3864148	GTG	TAA	2750	3037	96	Forward
3864172	GTG	TAG	311	862	184	Forward
3864180	~CAC	TTA~	930	1616	229	Reverse
3864184	GTG	TGA	197	670	158	Forward
3864184	GTG	TAA	612	1304	231	Forward
3864194	~CAC	CTA~	1084	1380	99	Reverse
3864338	GTG	TGA	552	1100	183	Forward
3864360	GTG	TAA	47	1078	344	Forward
3864388	GTG	TGA	1239	1586	116	Forward
3864406	~CAC	TTA~	263	958	232	Reverse
3864452	~CAC	TCA~	1079	1201	41	Reverse
3864458	GTG	TAA	797	1105	103	Forward
3864458	GTG	TGA	1179	1391	71	Forward
3864474	~CAC	CTA~	68	247	60	Reverse
3864474	~CAC	TTA~	644	1528	295	Reverse
3864510	~CAC	TTA~	1164	1640	159	Reverse
3864526	~CAC	TTA~	845	1660	272	Reverse
3864548	GTG	TGA	687	1055	123	Forward
3864548	GTG	TAA	979	1932	318	Forward
3864582	~CAC	TTA~	317	550	78	Reverse
3864604	~CAC	CTA~	1	141	47	Reverse
3864604	~CAC	CTA~	1513	1803	97	Reverse
3864610	GTG	TAA	427	1305	293	Forward
3864716	GTG	TAA	57	272	72	Forward
3864718	GTG	TGA	77	1474	466	Forward
3864802	~CAC	TTA~	92	550	153	Reverse

Assembly ID	Start	Stop	Start	Stop	Length	Direction
3864854	~CAC	CTA~	324	548	75	Reverse
3864862	~CAC	CTA~	431	1003	191	Reverse
3864888	~CAC	TTA~	10	657	216	Reverse
3864898	GTG	TAA	130	1029	300	Forward
3864938	GTG	TGA	883	1326	148	Forward
3864956	GTG	TAA	1030	1251	74	Forward
3864958	~CAC	TCA~	1427	1711	95	Reverse
3865022	~CAC	TCA~	279	1271	331	Reverse
3865036	GTG	TAG	79	492	138	Forward
3865054	~CAC	TCA~	302	793	164	Reverse
3865102	~CAC	CTA~	27	731	235	Reverse
3865156	~CAC	TTA~	416	808	131	Reverse
3865160	GTG	TAA	136	375	80	Forward
3865172	~CAC	TTA~	731	1123	131	Reverse
3865228	GTG	TAA	197	286	30	Forward
3865230	GTG	TGA	272	586	105	Forward
3865378	~CAC	TTA~	421	807	129	Reverse
3865470	GTG	TAG	98	742	215	Forward
3865632	GTG	TAA	46	456	137	Forward
3865710	~CAC	TCA~	287	448	54	Reverse

EXAMPLES

The examples below are carried out using standard techniques, which are well known and routine to those of skill in the art, except where otherwise described in detail. The examples are illustrative, but do not limit the invention.

Example 1

Isolation of DNA coding for a virulence gene in *Streptococcus pneumoniae*

As mentioned above each of the DNAs disclosed herein by virtue of the fact that it includes an intact open reading frame is useful to a greater or lesser extent as a screen for identifying antimicrobial compounds. A useful approach for selecting the preferred DNA sequences for screen development is evaluation by insertion-duplication mutagenesis. This system disclosed by Morrison et al., *J. Bacteriol.* 159:870 (1984), is applied as follows.

Briefly, random fragments of *Streptococcus pneumoniae*, strain 0100993 DNA are generated enzymatically (by restriction endonuclease digestion) or physically (by sonication based shearing) followed by gel fractionation and end repair employing T4 DNA

polymerase. It is preferred that the DNA fragments so produced are in the range of 200-400 base pairs, a size sufficient to ensure homologous recombination and to insure a representative library in *E.coli*. The fragments are then inserted into appropriately tagged plasmids as described in Hensel et al., Science 269: 400-403(1995). Although a number of plasmids can be used for this purpose, a particularly useful plasmid is pJDC9 described by Pearce et al., Mol. Microbiol. 9:1037 (1993) which carries the *erm* gene facilitating erythromycin selection in either *E. coli* or *S. pneumoniae* previously modified by incorporation of DNA sequence tags into one of the polylinker cloning sites. The tagged plasmids are introduced into the appropriate *S. pneumoniae* strain selected, *inter alia*, on the basis of serotype and virulence in a murine model of pneumococcal pneumonia.

It is appreciated that a seventeen amino acid competence factor exists (Havastein et al., Proc. Nat'l. Acad. Sci. USA 92:11140-44 (1995)) and may be usefully employed in this protocol to increase the transformation frequencies. A proportion of transformants are analysed to verify homologous integration and as a check on stability. Unwanted levels of reversion are minimized because the duplicated regions will be short (200-400 bp), however if significant reversion rates are encountered they may be modulated by maintaining antibiotic selection during the growth of the transformants in culture and/or during growth in the animal.

The *S. pneumoniae* transformants are pooled for inoculation into mice, *eg.*, Swiss and/or C57B1/6. Preliminary experiments are conducted to establish the optimum complexity of the pools and level of inoculum. A particularly useful model has been described by Veber et al. (J. Antimicrobiol. Chemother. 32:432 (1993)) in which 10^5 cfu inocula sizes are introduced by mouth to the trachea. Strain differences are observed with respect to onset of disease *e.g.*, 3-4 days for Swiss mice and 8-10 days for C57B1/6. Infection yields in the lungs approach 10^8 cfu/lung. IP administration is also possible when genes mediating blood stream infection are evaluated. Following optimization of parameters of the infection model, the mutant bank normally comprising several thousand strains is subjected to the virulence test. Mutants with attenuated virulence are identified by hybridization analysis using the labelled tags from the "input" and "recovered" pools as probes as described in Hensel et al., Science 269: 400-403(1995). *S. pneumoniae* DNA is colony blotted or dot blotted, DNA flanking the integrated plasmid is cloned by plasmid rescue in *E. coli* (Morrison et al., J. Bacteriol. 159:870 (1984)) and sequenced. Following sequencing, the DNA is compared to the nucleotide sequences given herein and the

appropriate ORF is identified and function confirmed for example by knock-out studies. Expression vectors providing the selected protein are prepared and the protein is configured in an appropriate screen for the identification of anti-microbial agents. Alternatively, genomic DNA libraries are probed with restriction fragments flanking the integrated plasmid to isolate full-length cloned virulence genes whose function can be confirmed by "knock-out" studies or other methods, which are then expressed and incorporated into a screen as described above.

SEQUENCE LISTING

(1) GENERAL INFORMATION

(i) APPLICANT: SmithKline Beecham Corporation and SmithKline Beecham p.l.c.

(ii) TITLE OF THE INVENTION: Novel Coding Sequences

(iii) NUMBER OF SEQUENCES: 185

(iv) CORRESPONDENCE ADDRESS:

- (A) ADDRESSEE: SmithKline Beecham Corporation
- (B) STREET: 709 Swedeland Road
- (C) CITY: King of Prussia
- (D) STATE: PA
- (E) COUNTRY: USA
- (F) ZIP: 19046

(v) COMPUTER READABLE FORM:

- (A) MEDIUM TYPE: Diskette
- (B) COMPUTER: IBM Compatible
- (C) OPERATING SYSTEM: DOS
- (D) SOFTWARE: FastSEQ for Windows Version 2.0

(vi) CURRENT APPLICATION DATA:

- (A) APPLICATION NUMBER: PCT/US97/19226
- (B) FILING DATE: 27-OCT-1998
- (C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:

- (A) APPLICATION NUMBER: 60/029,930
- (B) FILING DATE:

(viii) ATTORNEY/AGENT INFORMATION:

- (A) NAME: Gimmi, Edward R
- (B) REGISTRATION NUMBER: 38,891
- (C) REFERENCE/DOCKET NUMBER: P50577

(ix) TELECOMMUNICATION INFORMATION:

- (A) TELEPHONE: 610-270-4478
- (B) TELEFAX: 610-270-5090
- (C) TELEX:

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 495 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

CTCGGTGATA GAAATAGTGT AATCATGCTT TTCTCTTCTT ATCTATACTT TGCTACTTCT	60
ATTATACAAA AAAATAAAGC GCTTGACTAG GGATTTTGTAG AAAAAAAGCC TATTTTTCAT	120
AGAAAAATAG GCTTTTTCG AACGATTGAC ACAATTGGAT TTGGTTAATT CACTCTTAAC	180
GATGGTTTTTA AACGATATAT ATTTTATATAT ATGTAAATTA AAAACTTCTT TCCTTTCAC	240
TCCTACGACT TTTCAGATAC AGATAGCCAA AGAAGTTTTT ATAGAGGGCA AAAAAAGAGGA	300
GGAAGGCATG AAGAAAGAAG GTCTCTGGCA AAATCATAAT AACAGGATCC TTGGCTGGAT	360
CAAAAAGCCA GGTATCATCT CCCACAAAGA GAATTTGATG GAAAAGAGTA AAGAATTGGT	420
CAAAACCAAT CAAAACCTCC CCAAGTCCAT CATCACAGGT AAGACTACTA GAGCCAGGAG	480
ACTTTTTCGA TAAAG	495

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 529 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

CTAGAGCAAG TATTTTTCAT ACTTTTTCCT AATAAATAGA TAGAGCCAGA GAATTTAGTA	60
AACCTAGATT TAAAAATGTG CTATAACATA ATATATTGAA TCTATAATAG TACACCTTGA	120
CTGCTAAAT ATTTCTATAA ATTAATTTGA CTTTCCTGAT AGAGTTATTC ACATCTTATT	180
TCAACTCACT ATAGAAGGAG GAATAGGAGG ATTCTCAGAC ATCCGGGCAT CAGCCCAACT	240
AATGATTTGA TTGCTAAGAA AATATTTCAGC AATCCAGAAA TCACTTGTCAT ATTTATTTCG	300
GATATGCTGG ACTTGCCAGC AAAAAATGTT GACCATTTTG GAGGGAAGCG ATATTCACGT	360
ATTACTCTCC ATGCCTTACT CAGTGCAGGA TTTTATATACC AGTATAGACG TCTTGCCGGA	420

GTTGGATAAC	GGTACTCAAG	TAATTATTGA	GATTCAAGTC	CATCATCAGA	ATTTTTCATC	480
AATCACTTGT	GGACTTACCT	GTGCAGTCAG	GTTAATCAAA	TCTTGAAAA		529

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 885 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

CTCATCATCT	GTCAAAAAGC	GTTTCTTAGC	AGTCGTGATA	TCCATAAAAT	AATCTAATAT	60
CACGATTTCC	TCATCCGCAA	AGAAAGGAAG	GCTGACCAAC	TCCAGTGCCA	CATCCTTGTA	120
AACTACTTCT	TGCATATCAA	AGTAGGCAAA	GTTGAGGTCA	GCAGAATCAT	ACCCAATCTG	180
TTTCAACACT	TGACTCTTCA	TCACTTCAAA	CTGACCCTGA	TCTGTCCCTG	TAAATAGGCG	240
CAGGCTCGGT	AAATTCGATA	AAGTCAACTT	CTGACTTTCT	TCAATGGCTA	GCATCGTCTC	300
TCCTTTCTTC	AGATTTTTTCG	ATTTAATTTA	GTCAATATAG	CGCAATTTCC	CACGGAAATC	360
TTCTAAGCTC	TCGTAGCCTT	TTTCCACCAT	GATTGCTTTC	AGTTCATTGG	TAAAGCGGTC	420
AAAAGCACTG	ACGCCTTCTT	TGTGAAGGGT	CGTTCCCACC	TGCACCATAC	TTGCTCCACA	480
GAGGATGTGT	TCAAAGGCAT	CTCGACCAGT	CAGAACGCCA	CCTGTTCCGA	TAATTTGGAT	540
TTGAGGATTT	AAACGTTGAT	AAAAGGCGTG	AACATTGGCT	AGAGCAGTCG	GTTTGATGTA	600
TTATCCACCA	ATTCCACCAA	AACCATTCTT	AGGCCGAATA	ACGACAGATT	CGTCTTCTAT	660
ATAGAGGCCG	TTTCCGATAG	AGTTAACGCA	GTTGACAAAC	TTGAGCGGAT	ATTTGTTGAA	720
AATAGCTGCC	GCTTGATCAA	AGTGAACAAT	ATCAAAATAA	GGTGGCAATT	TAATTCCAAG	780
AGGTTTGGTG	AAGTAAGCAA	ACACTTCTGC	CAAAATCCGG	TCTGTTGTCT	CAAAATCATA	840
GGCAATCTGA	GGTTTACCTG	GAACATTTGG	ACAGGAAAGA	TTTAG		885

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 925 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

TCTTGGCCAA	CTGCATGGAG	TTCAGCGGTC	AATTTCAACG	CACCTGAGAA	ACAGACCCCT	60
GCACCCCTGA	AATCTCAGGA	GACATGATGG	TCTGGATGGA	ATCAATAATG	AGAAAGTCTG	120

GCTGGATACG	CTACCACTTC	TGCACGAACA	CTCTGCATAT	TGGTCTCTGC	ATAGAGATAA	180
AACTCACTAT	CAAAATCACC	TAAGCGCTCT	GCACGTAGTT	TAATCTGCTG	GGCAGACTCC	240
TCCCCACTGA	CATAGAGAAC	TGTCCCCACT	TGGGACAAC	GGGTTGAGAC	TTGTAGGAGA	300
AGAGTTGATT	TCCCAATCCC	AGGATCCCCA	CCGATGAGGA	CGAGACTTTC	CTGGTACAAC	360
TCCGCCTCCA	AGCACACGGT	TGAATTCCTC	CATCTCCGTC	TTGGTTCGAT	TGACATTGAT	420
GGAAGTCACC	TCAGCTAGTT	TCATGGGCTT	GGTTTTCTCA	CCTGTCAAGG	ACACACGCGC	480
ATTCTTGACC	TCGGCAACCT	CAACCTCTTC	CACAAAAGAA	GACCAAGACC	CACAGTTGGG	540
GCAACGTCCC	AGATATTTAG	GGGAATTATA	CCCACAATTT	TGACATACAA	ATGTCGCTTT	600
TTTCTTTGCG	ATGACAAACC	TCTTCTATA	TCTCTAATC	ACACTCAATC	ACTTGGCAAA	660
AATCAATCTT	CTCATTTGGC	ACAAACTGGC	GCATGAGCAT	TCGATGAGCA	ACAACTACCA	720
CAGTCTGATG	TTCTCGATAC	TTAGACATAC	ATTCTAGAAA	CCGAGACTTC	ATTTCCGTAG	780
CTGTCTCATA	TTGAATAGGA	CTATTAGGAA	GCAACTCCCC	CTTGTTTTCT	AAAAACAGTC	840
TTCTAGCTGT	TTCAAAGTTT	TCTATTCCTG	TTTTATAGAC	CTGCCATTCA	TGTAATAAAG	900
GCTCTACTCT	TAAAGGAAGA	CCCGT				925

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 602 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

TTATGTCAGT	GGGATTACGC	CTAATCTCCC	AGAAGCAGAA	TTATTATCCG	GTCAGGAAAT	60
TAAAACCTTG	GNAGACATGA	AAACTGCAGC	GCAGAAATTG	CATGATTTAG	GAGCGCCAGC	120
AGTCATTATC	AAAGGGAGGC	AATCGTCTTA	GTCAGGACAA	GGCTGTGGAT	GTCTTTTATG	180
ATGGACAGAC	CTTTACTATC	CTAGAAAATC	CAGTTATCCA	AGGCCAAAAT	GCTGGTGCAG	240
GTTGTACCTT	TGCCTCTAGC	ATTGCCAGTC	ACTTGGTTAA	AGGTGATAAA	CTTTTGCCAG	300
CAGTAGAAAG	CTCTAAGGCT	TTCGTTTATC	GTGCTATTGC	ACAAGCAGAT	CAGTATGGAG	360
TAAGACAATA	TGAAGCAAAC	AAAAACAAC	AAAATCGCCC	TTGTATCCCT	ATTAACCGCC	420
CTTTCTGTGG	TTCTAGGTTA	TTTCTTAAAA	ATCCCAACAC	CTACAGGNAT	TCTAACTCTT	480
TTAGATGCTG	GTGTCTTCTT	TGCGGCCTTT	TACTTTGGTA	GTCGTGAAGG	AGCGGTAGTC	540
GGAGGACTAG	CAAGTTTCTT	GCTTGACCTC	TTATCAGGCT	ACCCTCAGTG	GATGTTTTTT	600
AG						602

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 456 base pairs
- (B) TYPE: nucleic acid

- (C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

CTGGATACTA	AGAGAAATCA	AAAAAGCACT	CTAGGATAGA	GGCCTAAAGT	GCTTAGTTTC	60
AAGGCTTTAC	AGCCTATCAT	ATTTAATAAA	ATATTACAAC	ATCTTGTTGT	AGAATTCAAC	120
GACAAGTGCT	TCGTTGATTT	CTGGGTTGAT	TTCTGTCGCGT	TCTGGCAAGC	GAGTCAATGA	180
ACCTTCCAAT	TTTTCAGCGT	CGAATGATAC	GAATGCTGGA	CGTCCAAGAG	TAGCTTCTAC	240
TGCTTCAAGG	ATTGCTGGAA	CTTTCATGA	TTTTTCACGA	ACTGAGATCA	CTTGACCTGC	300
AGTTACGCGG	TATGATGGGA	TATCAACGCG	TTTCCCGTCA	ACAAGGATGT	GACCGCTGGT	360
TTACAAATTG	GACCAACTT	GACGACCAGT	AGTCGCGAGA	CCAAGACGGT	AAACAACGTT	420
ATCCAAACGA	CGTTCCAAA	GAAGCATAAA	GTTGAA			456

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1961 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

CTAATATAGA	ATAATCACCG	CCGTTGTGAA	AGAACGATTG	GATGATAATC	CAATCGTTCA	60
GGGAAATTGG	AAGACCTTGG	GTTTCCAATT	TAGGCATGAG	ACACCTTTGG	TGGCTGCTGC	120
CGTCCCTCAC	AAGCTAAGGT	GATTGTTGAA	AAAGAGGAAA	AAGGAGAAGA	AATGAAACCA	180
GTAATTTCCA	TCATCATGGG	CTCAAAATCC	GACTGGGCAA	CCATGC AAAA	AACAGCAGAA	240
GTCCTAGACC	GCTTCGGTGT	AGCCTACGAA	AAGAAAGTTG	TTTCCGCACA	CCGTACACCA	300
GACCTCATGT	TCAAACATGC	AGAAGAAGCC	CGTAGTCGTG	GCATCAAGAT	CATCATCGCA	360
GGTGCTGGTG	GCGCAGCGCA	TTTGCCAGGC	ATGGTAGCTG	CCAAAACAAC	CCTTCCAGTC	420
ATTGGTGTGC	CAGTCAAGTC	TCGTGCTCTT	AGTGGAGTGG	ATTCACTCTA	TTCTATCGTT	480
CAGATGCCGG	GTGGGGTGCC	TGTTGCGACC	ATGGCTATCG	GTGAACTCTT	TTTTAGGATA	540
TAAAACAGGG	TTCCGATAAG	TTTTTTTGCA	AGGTGGATGA	TGGCTACATT	GTAATGTTTT	600
CCTTGTTCTA	ACTTAGTCTT	AAAAGCAGGT	GAAAAGTGAG	GGCATGCTTT	GGCAGCTTGT	660
ATGAGTACCT	ACCGCAGATA	AGGGGAACCC	CGTTTGACCA	TCCTCCCAGC	TAAATCAATC	720
TGACCTGACT	GATAAATAGA	AGAATCCAGT	CCAGCGAAAG	CTTGTAATTG	AGCAGGATTA	780
TCAAAGGCAT	GAATATTTTCG	AATCTCGGCT	AAAATGACCG	CCCCTAAACG	ATTCTCAATC	840
CCAGTAACCG	TCGTGATGAC	CGAGTTTAAC	TCAGCCATCA	AGTCATTGAC	ACATTTTTTC	900
GCCTTGTC	TGAGCCTCTT	GTAATGTTTG	ATGTTTTTCAT	TACACGAGAT	AAAACGTCTA	960
TGCGTTATCA	AACTCATTAC	CAATTAAAC	AAATGTGGTT	AGATCCTTTC	GGAAATTGTC	1020

AAGCGATTGG	AGGAAATGAA	CTAATCCACA	GCGGCTTATT	CCAAGTATAC	CACTTGGGCT	1080
TTGGCAGTAG	CTAACTGCGC	TAAATATAAT	ATAAGGAGGA	GTAAAATGAA	GACAGTTCAA	1140
TTTTTTTGGC	ATTATTTTAA	GGTCTACAAG	TTCTCATTTG	TAGTTGTCAT	CCTGATGATT	1200
GTTCTGGCGA	CTTTTGCCCA	AGCCCTCTTT	CCAGTCTTTT	CTGGACAAGC	GGTGACGCAG	1260
CTAGCCAATT	TAGTTCAAGC	TTATCAAAAT	GGGCAATCCA	GAACTTGTAT	GGCAAAGCCT	1320
ATCAGGAATT	CATGGTCAAT	CTTGGCCTGC	TGGTTTTGGG	TTCTATTTAT	CTCTAGGTGT	1380
AATATAAACA	TGTGTCTCAT	GACGCGCGTG	ATTGCAGAAT	CGACCAACGA	GATGCGCAAA	1440
GGTCTCTTTG	GTAAGCTTGC	TCAGTTGACG	GTTTCTTTCT	TTGACCGTCG	ACAAGATGGC	1500
GATATCCTGT	CTCATTTTAC	CAGTGATTTG	GATAATATCC	TCCAAGCCTT	TAACGAAAGC	1560
TTGATTCAGG	TCATGAGCAA	TATTGTTTTA	TACATTGGTC	TGATTCTTGT	CATGTTTTCG	1620
AGAAATGTGA	CGCTGGCTCT	CATCACCATT	GCCAGCACCC	CATTGGCTTT	CCTTATGCTG	1680
ATTTTCATCG	TGAAAATGGC	ACGTAAATAC	ACCAACCTCC	AGCAGAAAGA	GGTAGGGAAG	1740
CTCAACGCCT	ATATGGATGA	GAGCATCTCA	GGCCAAAAAG	CCGTGATTGT	GCTAGGAATT	1800
CAAGAGGATA	TGATGGCAGG	ATTTCTTGAA	CAAAATGAGC	GCGTGCGCAA	GGCAACCTTT	1860
AAAGGAAGAA	TGTTCTCAGG	AATTCTTTTC	CCTGTCATGA	ATGGGATGAG	CCTGATTAAT	1920
ACAGCCATCG	TCATCTTTGC	TGGTTCGGCT	GTA CTTTGA	A		1961

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 375 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

CTATCTCCAA	GTNCGNTTGG	AATNCCTCCG	CNANCCACAA	CTCATCCAAG	CACTTTNCAA	60
CGTGNCCTGG	TCCGGTCCTC	CAGTGCGTCT	NACNGCACCT	TCAACCTGCN	CATGGGTAGG	120
TCACATGGCT	TCGGGTCTAC	GTCATGATAC	TAAGGCGCCC	TATTCAGACT	CGGNTNCCCT	180
AGGGCTCCGT	CTCTTCAACT	TAACCACGCA	ACAGAACGTN	ACCCGCCGGT	TCATTCTACA	240
AAAGGCAGNC	TCTCACCCAT	TAACGGGCTC	GAAGTTGTTG	TAGGCACACN	GCTTCAGGTN	300
CTATTTTACC	CCCCTCCCGG	GGAGCANCTC	AACTGACCCN	CACGGCACCG	GTGNANNAAA	360
CGGTCACTTA	GGGAG					375

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 665 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

GGGGGGGGTN	NNTTCTGGGG	CCGGGTGNNT	CCTNGAAAAA	ATGCTGGACT	TAACGGTTAA	60
ATCATTTGAA	TTGGCCTGTG	GATTTTAGCT	AGCAATCCAG	AGCGAGTTTT	CTCCAAGACA	120
GACCTCTATG	AAAAGATCTG	GAAAGAANAC	TACGTGGATG	ACACCAATAC	CTTGAATGTG	180
CATATCCATG	CTCTTCGACA	GGAGCTGGCA	AAATATAGTA	GTGACCAAAC	GCCCACTATT	240
AAGACAGTTT	GGGGGTGGG	ATATAAGATA	GAGAAACCGA	GAGGACAAAC	ATGAAACTAA	300
AAAGTTATAT	TTTGGTTGGA	TATATTATTT	CAACCTCTTT	AACCATTTTG	GTTGTTTTTT	360
GGGCTGTTCA	AAAAATGCTG	ATTGCGAAAG	GCGAGATTTA	CTTTTTCGCT	GGGATGACCA	420
TCGTTGCCAG	CCTTGTCGGT	GCTGGGATTA	GTCTCTTTCT	CCTATTGCCA	GTCTTTACGT	480
CGTTGGGCAA	ACTCAAGGAG	CATGCCAAGC	GGGTAGCGGC	CAAGGATTTT	CCTCCAATTT	540
GGANGTTCAA	GGTCCCTGTT	AAATTTCCCC	CATTTAGGGG	CAACCTTTTA	ATGAAANTTT	600
CCNTNATTTG	CCGGGTANCT	TTGAATCCCT	NGGAAAAAAC	CCAACNAAAA	AAAGGGCTTA	660
NNCCC						665

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 989 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

CTACGATATC	TTTGGTCTTT	TGTAAGATAT	GAGGTCCACC	CTTATGCGCC	TCAGTTGGCA	60
TTTCATGCGA	TTCAAGAAGT	TGCCCCCTCT	GATCAACCAA	ACCATACTTG	ATGTTGGTTC	120
CACCGATATC	AATTGCAACG	TAATATGTCA	TAAATACCTC	CTTTTAGATT	AGAGGAAGCG	180
CTCCTTGGTT	TCACGAATCA	AGGCAGCAGC	CGCTTCTACA	ACTGGACGAT	CTTCTTCAGT	240
CACTGGTGTC	AATGGTGAAC	GAACAGATCC	AATATTCAAG	CCTTCATTGA	TTTTCAAGAC	300
TTCTTTGATG	ACACCGTACA	TATTTCCATG	AGCAGAAGTG	AGTTTACCAA	TGATTGCGTT	360
GATAGCATAC	TGCAATTCAC	GCGCTGTTTC	TAGGTCCCTA	TCCGCAATCA	ACTGATTGAG	420
TTTCAAGAAG	AGTTCTGGCA	TAGCACCATA	AGTACCACCG	ATACCAGCCC	TAGCCCCCAT	480
GAGGCGTCC	CCTAGGAAC	GCTCATCAGG	ACCATTAAAG	ACGATATGGT	CTTCTCCACC	540
AAGGCTGACA	AAGGTTTGGA	TATCTTGAAC	TGGCATAGAA	GAGTTCTTCA	CACCGATAAC	600
ACGAGGATTT	TTCAACATTT	CTGTGTAAAG	GCTTGGAGTC	AAAGCAACCC	CTGCCAATTG	660
AGGAATGTTG	TAAATCACGT	AGTCTGTGTT	TGGAGCTGCA	GAAGTGATAT	CGTTCCAGTA	720
TTTGGCAACT	GAGTTATTCT	GGCAAGCGGA	AATAAATTGG	TGGAATCCGT	TGCAATAGCA	780
TCTACTCCCA	AGCTTTCAGC	ATGGCGAGCA	AGTTCCATAC	TATCTTTAGT	ATTATTGCAA	840

GCAACATGGG	CAATAATGGT	CAATTTACCT	TTGGCTACCG	CCATGACTTC	TTCCAAAATC	900
AACTTGCGAT	CTTCAACGCT	TTGGTAGATA	CATTCACCAG	AAGAACCATT	GACATAAGAC	960
CTTGAACACC	TTTATCAATG	AAGTATTGA				989

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1450 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

CTCCATATTT	CTTAGCCTTC	TCAATTAGGG	TCTTGAAGTC	TTCGACACCA	CCGATACGCT	60
TACCAATATC	AGCATAGTTC	AAGTGACCAG	AGTCATGGCT	GTGATATCCT	TAACTTTTTC	120
CCAACCTTGA	GGGTTGTTCA	TAATGCTACG	ATAAGCAATG	GCACCATCTT	GCCAATCAAC	180
TTTCTTGTCT	GCATTGGCAT	CTTCAGTGAT	AACAACCTTA	GCACTTGGAA	GTTCTTCGT	240
GTATTCTGGG	AAAACAATGC	CCTTATAAGC	TTTTTCCCAT	TGCCATTTCAG	AGCTGTGGAT	300
TCCTACATAG	TTGGCATTTT	CGACTGTTTC	TTTATAAGCT	GTCAAACGAG	TCCAGTCATT	360
CGAACCACCA	CCATAGCTAT	TTTGAGAGTT	ACTCCAAACA	CCAGCAGCAA	GCTTATCTGT	420
AGAAACAAAT	CCATACATGT	AACCCTTAGC	CAAAATCCTT	ATTGGATTGG	TTACATCGAT	480
ATGATCATCT	CCGCTGACAT	GCGTATTGTT	TGACATGGTT	GCCCCATCAA	ACTTAGCACC	540
AGTTTGATCA	CTAGAAACAG	AGACTAAAGC	ATTGCCGAGG	AAACTAATAG	AAGAAAGTAG	600
TTTTCTTTTCG	TCATCAATCT	TTTGACCTGG	AGTGACTTGA	TTGTGGTTGA	CAATCTTGGT	660
CACATCAAAG	TGCAATTGAT	TGTCCACAAC	TTGCAAGCGT	ACTGTCATTT	CCGCATTGAT	720
TAAGTGAGCA	TCATCGCGAA	GCTTCATCAA	GTACTCTGCT	GTTGTCTCAT	TGATTTTTTTT	780
ATAAGTGACT	TCAGGGGTGA	TTCGGTGGTT	ATTGATAAAG	ACTTGGTTGA	ATTGTTGCAC	840
CTGTCCCTGGC	AAAGTATGTC	CATTCAAGGT	GTATCCCTTG	ACACGAAGGA	AGGCTTGGTC	900
AATTACTGCC	TTAAGTACCT	TAAACTGGAT	CGTATCATAA	GTCACCTTGC	TATCGTCAAC	960
AACCGGACCT	GTTTCTTTCT	GGGCAGGGGT	ATCCTCTGGG	TTTTACCCTC	TCTGTGGCTA	1020
TCCGTTTCAA	CGCTTGAACA	ACTGGTCGCT	CATCGTCATA	AGAGCCCGCC	TTGAGAAAAA	1080
TCTTCTTCTC	ATTTCTAAGA	TGGTCATTGA	CCGCAGCTGG	TAGAGTCACT	GTGTCAAAGA	1140
AGATTGACAT	CCTTATTTGC	CTGGCATTTA	CCTGACCGTC	TGACTTGAAG	ACTGATAGAG	1200
AGACGGTTTG	TTGATCCTGT	TTCAGGAGCA	GCAACACGAC	TACCTCTATA	CCAAGTGCTA	1260
GTTGTTGGAG	ATTTATACTC	CCAGAACCAG	CCATCCTTGT	CATAACCGAC	AAAAACATTA	1320
TTATTGGTAT	CTTTAAATTT	CAAGGAGACA	CCAAAGCGTG	ATTTGCCCTT	TTCAGAATCT	1380
TCTTTGAAGG	TTAAATCAAC	AGTTGCATTT	CCATTGGCAT	CAACGGTCAA	GCCCTTCTTT	1440
TCAAACAGAG						1450

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 420 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

CTGCGAGTTG	TGAGGCTCCT	ATTATGTCTC	GTGATTAAAA	TCTCTATAAG	GTGATTTTGG	60
AGGGAAATTA	TCGGGCGACA	GCGGGTAGAG	AAGAGATGAA	AGAGGCTATT	TTGGAATATC	120
AAGCAAATCC	TGCTGCCTTA	AAAGATCTCA	AAGAAAAGGC	TAAGAATATT	TCCAGAGAGT	180
ATTCTGAAGA	GCATCTGTTA	CAAATCTGGT	TGGACTTTTA	TGAGAAACAA	GCCGCTTTAG	240
GGACAAAGTA	AAAAGTGAGG	TAATCTATGC	GAATTGTTTT	ATTTACAGAT	ACCTATTTTC	300
CTCAGGTTTC	TGGTGTTCGC	ACCAATATCC	CAACCTTGAA	AACCCACCTT	GAAAACACGG	360
ACTTGCCTGC	ATTTNTATCT	CATACAATCC	ACCGAATTTC	GATGTCCCCC	TCCCTACAAC	420

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 661 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

CTCCCCAAAC	TTTTATTTGA	GAGTGAACGG	TATAAGAATA	TGAAACCGGA	GGTTAAGGTG	60
GTTTACTCAG	TTTTAAAAGA	TCGGTTGGAG	TTGTCTTTGA	GCAAAGGTTG	GATTGATGAG	120
GATGGGACTA	TTTATTTGAT	TTATTCCAAT	TCAAATTTGA	TGGCACTTTT	AGGCTGTTCA	180
AAGTCAAAAT	TACTCTCCAT	GTGAGTTTGA	AGTGACATTT	TTAGATGATT	ACCATAAAAA	240
ACATAACTAC	CCACTATTTT	ACGAATCCTA	TCTTCAAAAC	GTTATGGAAT	TCCTTGAAAG	300
TCAAGACATA	AAGAATGGGG	TTGATGCCTT	TGTAGATGAT	CATCAAAATC	TCGTTTTTGT	360
TTTATATGGA	CAAGGCTATC	GAGCCGAGGG	AAAAGAGGGA	ATACTTACAA	CCCAAGTAAC	420
TGTAAAAGCT	TATGATGAAG	ACAAGAAACC	GATTAACTTC	GCAAATTTAT	TAGATTCCTT	480
AATCGTGTCA	GAATATCAAA	TGGAACCGAA	TCTTTGGGAG	GTCTCCTATG	ATTGATCTCT	540
ATCTAAGTAA	AAATAGCCGA	AGAAATCAAC	TTCTTTTAGA	CTTCTTCCAA	AACTATGGCA	600
TCGAGGTATC	TTGTCATTCA	GTTTCTGAAA	TGACAAAGGA	CAAATTAATT	GAGATGATGA	660
G						661

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1429 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

CTGCCCCGTGT	AAGGCTGGAC	GATTGCCTTT	CTTAGTATCC	GCAAAGAGGT	AAACTGAGAA	60
TAGAGAGGAT	TTCTCCTTCA	ATATCTTTGA	CAGACAGGTT	CATCTTGCCT	TCTACGTCTG	120
AAAAAATCCG	CATATTGACC	AGTTTTCTCA	CAGCATAGTC	CAAATCTTCC	TCTTGGTCCT	180
CTGGTCCAAC	ACCAACCAGC	AATAAAAGTC	CCTGATTGAT	TTTTCCCTGA	ATCTGGCCTT	240
CTATACTCAC	TTGGGCTTTT	TTAACCCGTT	GGATAATGAT	TTTCATAATA	GCCTTTCTAG	300
TAAGAGCTAG	GACAAC TAGC	CGTTGGTCCG	TTTGACAGAG	TAAACTTCTG	GCACACTCTT	360
AATTTTATCG	ACAACCGTGG	TCAGTGTAGA	GAGGTTGGCA	ATACCGAAGG	ACACATGGAT	420
ATTAGCAAAC	TTCATATCCT	TGGTTGGTTG	GGCATTGACC	GTTGAAATAT	TCTTGGTTGT	480
ATTTGAAAGA	ACTTGACAGTA	CATCGTTCAA	CAGTCCTGTA	CGGTTGAGAC	CGTAGATATC	540
GATATGGGCC	ATATACTCCT	TATTTGAGCT	AGAGTACTGG	TCTTCCCATT	CCACATCAAG	600
GAGACGTTGC	TCGTAGTTTT	CTTGGGCACG	CAGGTTTATA	CAGTCCACAC	GGTGAATAGC	660
CACACCACGA	CCCTTGGTAA	TGTAGCCAAC	AATATCGTCA	CCAGGCACGG	GGTTACAACA	720
CTTAGCAATC	CGCACTAGGA	GACCAGAAGC	ACCTTCAATA	ACCACTCCCC	CCTCATGCTT	780
GACCTTGAG	AGTTTCTTTA	TTTTCAACCT	TGACCTCGCC	ACCTTTGACA	AGCTCCTCTG	840
CCTCAGCCTT	GGCCTTGGA	CGCTCTTCCT	CACGGCGTTC	TTTTTTCAGTC	AGACGGTTAA	900
AGACGGTAAT	CGCACCGATT	TCCCCAAAAC	CAATGGCCGC	AAAGAGGGAG	TCTTCTGTCT	960
TGTAAC TGGT	CTTTTGCAGA	ACTTGATCCA	TGTGGCGCTT	GTCCATAAAT	TTATTTGCCA	1020
CATAGCCATT	TTCTTGGAAC	TGAGCCATCA	GCATCTCACG	ACCCTTGTTG	ACAGACAATT	1080
CCTTATCTTG	GTTTTTAAAG	AACTGGCGAA	TCTTATTGCG	CGCCTTGCTA	GTCTTGACCA	1140
TATTGAGCCA	GTCACGGCTA	GGTCCAAAGG	AGTTCGGGTT	GGCGATAATT	TCAACCTGAT	1200
CCCCTGTCTT	TAACTTGATT	GTCAGTGGAA	CCATGCGGCC	ATTGACCTTG	GCACCAGTTG	1260
CTTTTTTACC	GACCTTGATA	TGGATTTTCGT	AGGCAAAATC	AATCGGTCCT	GAATCTTTGG	1320
GAAGAGAACG	GACAGCTCCA	TCTGGGGTAA	AAACGTAAAT	CTCCTCAGCC	AGATAGTTTTT	1380
CCTTAACAGA	GTCCACAAAT	TCCTTAGCAT	CATCAGCCTG	GTCTTGAG		1429

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1513 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

CTCTGCAATG	ATGTACTCAA	ACATCTCCGC	TTCTAGTTCC	TCCTTAGGCA	GAGGCAATTT	60
CCCACGTCGC	ATCCGGTTCA	TAAAGACCGT	ATGGTTTCT	AAAATCAAAC	TATACAAACT	120
CATGTGGGGA	ATATCCAATC	CAATGGCTTT	AGCCACATTT	TCCTTTACTT	GCTCCATGGT	180
CTGACCAGGC	AGAGCATAAA	TCAAATCAAT	GGAGATGTTG	TCAAAACCAG	CCAGTTTCAG	240
GCGATCGATA	TTTTCATAAA	TATCCTTCTC	CAAATGACTG	CGCCCAATCT	TTTTCAACAT	300
CTTATCATCA	AAGGTCTGGA	CACCTAGCGA	AACACGATTG	ACAGCCGAAT	TTTTCAAAAC	360
AGCTATCTTA	TCCGCATCCA	AATCGCCTGG	ATTGGCTTCA	ATGGTCAACT	CTTCCAAGAC	420
AGACAAATCC	AAGTTTTTAG	TCAAGCCATT	CAGTAACACC	TCCAGTTGCG	GAGCCGACAG	480
GGCTGTCGGT	GTTCCACCAC	CGATATAAAG	GGTTGACAAC	TTTTCAATAT	CATAAGAACG	540
AAACTCTTCC	AGCAGATGCT	CTAAATAGCT	GTCGACTGGC	TGATTTTTGA	TGAAGACCTT	600
TGAAAAATCA	CAATAATAAC	AAATCTGGGT	ACAAAATGGG	ATGTGCACAT	AGGCTGACGT	660
TGGTTTTTTC	TGCATAGTAA	TTATTATACC	ACAAAGACTA	GATTCCAGAT	AAAAATCACC	720
ATCCCCAGAT	ACATAGTCCG	TCCGGAGATG	GTGATGGTTT	ATTCTTCTGT	TATATCAATC	780
ACAATCTCTT	CTGAGTCATC	AAGAGCTTCG	GCTTTTTCTT	GCCATTGTTC	CTTGAGATTA	840
TTTAATTGAT	TTTTTGATGC	TTCTGTCGCT	TGAAAAGCAT	AGGATTTAGC	TTGAGCAAGT	900
ATACTGTCCA	CAGTGATTTC	ACCTGACTCA	ACCTGTTCTT	TTGTTTTTCAG	AACAAAATCT	960
GTAGCCTGCT	CCTTAACCTC	TGTCAGTTTT	TCACAGACTT	GCTCCTTGGC	ATACTCCGGA	1020
TCTTCTCTCA	AATCATCTAA	AAAATCTTGA	GCCTGACTGC	AACTTGTTTT	GCCCTTATCA	1080
CTTGTTAAAA	ACAAGGCAAG	AGCTGCACCT	GAAACGGTTC	CTAAAAGGAT	TGAGGATAAT	1140
TTACCCATAA	GGATTCTCCT	TTTTTATTTT	TTGAAAAATT	TACTTGCAAG	ACGAAGAGCT	1200
GACAGACTTG	CACCAGTCTT	GAGTGTTTTT	GAACCAGCTG	ATGAAGCTTT	CTTGCTCAAG	1260
ACACGCGCAT	GGTCATTGAG	GTCTGAAACA	GATAGAGATA	AATCTGCAAC	AGCACTGAAG	1320
AGTGGATCAA	TCGTAGCCAC	CTTGACATTG	ATATCATCTG	CCAAGACATT	GACCTTAGCC	1380
AACAATCAT	TGGTGTGATG	CAAGGTCACA	TCCACATCTG	AAGTCAAGGT	TTTAATCGTC	1440
TTTTCTGTTT	CATCGATGAC	ACGACCAAGC	TTTTGTACAG	TAATGATCAG	ATAGACCAAA	1500
AAGACAATCA	CAG					1513

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 505 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

CTCTTGTCAG	AGAAATTTAC	AAAACGTTAG	GAGAATAAGA	TGGCATTAT	TGAAAAGGT	60
CAAGAAATCG	ATATGGAAGT	CATCAAGGCT	GAAACCCAAT	TGTCTGCAGA	AGCCTTGAGA	120
CTCAAGGAAA	GCCGTGACAG	GGAATTGGCA	GATATTATTT	CAGGGGAAGA	TGACCGTATT	180

CTCTTGGCTG	ATTGGTCCTT	GCTCTTCTGA	TAATGAAGAG	GCGGTCTTGG	AATATGCTCG	240
CCGTTTATCC	GCCTTGCAAA	AGAAGGTAGC	GGATAAGATT	TTCATGGTCA	TGCGCGTGTA	300
TACTGCTAAG	CCTCGTACCA	ATGGAGACGG	CTATAAAGGG	TTGGTTCACC	AGCCAGATAC	360
TTCTAAGGCT	CCAACCCTGA	TTAACGGCTT	GCAGGCTGTG	CGCCAGTTGC	ACTACCGCGT	420
TGATTACAGA	GACTGGTTTG	ACAACGGCAG	ATGAGATGCT	TTATCCGTCA	AATCTGATCT	480
TGGTGGATGA	CTTTGGTCAC	CTACC				505

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1827 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

CTCTTTTAAAC	CGTTTTAGCG	GTGACACCGA	GGATTTTTTTC	AGGACCCAAG	ACTTGTCGGG	60
CAACCGAAAC	TGGGAGTTCG	TCATCTCCAA	TATGCAGACC	AGCAGCATCA	ACCGCAAGAC	120
AAACATCCAA	CCGATCATCG	ATTATCAAGG	GGACCTGATA	GGCATCTGTT	ATTTCCCTTGA	180
CTTGTTTTTGC	CAGTTGATAA	TATTGATTGG	TTGTGAGATT	TTTTTCTCGC	AATTGGACTA	240
TGGTAACCCC	TGAACGGCAG	GCCGTCTCAA	CTTTTGCAAG	AAAGCTTTCC	ACGGAATCTT	300
GATAGCGATT	GGTTACCAGA	TATAGTCTAA	GCGCTTCTCT	ATTCATAAAC	CTCTCCTTTG	360
ATGGTATCTA	GCCAATTTTC	ATCTCTTCTT	AGGAGCGAAA	GCTGATTGAG	TACTTGGTAA	420
CGAAATTCTT	CCAATCCCAT	TCCTTGAACA	ACTATTTTCT	CAGCAGCGAT	ATTGAGATAA	480
GAGACTGCTA	AGCAAGAACT	TCAAAACCAG	TCTTTCCTTG	GCTGAGAAAA	ACAGCTGTTA	540
AGGCTCCAAC	CAAGTCTCCT	GTCCCTGTTA	TCCAGTCTAA	TTCAGTACAG	CCATTCTCAA	600
GTACAGCAAC	TTGATTCTCC	GAAACAATAA	GGTCCTTGGG	ACCTGTGACT	AAGAATGACA	660
TACCACGATA	GGTCTGACAC	CAGTCTTTCA	AGACTTGAAG	CAAATCCTCC	GTTTCTTGAT	720
CTTTAGCACT	CGCATCGACC	CCAACGCCGT	GATGCTTTAA	TCCAACAAGA	CTTCGAATTT	780
CTGACATGTT	TCCTTTAAGG	ACCGTAGGTC	TATAGTCTAA	AAGGTCTTTA	ACTAAGCTCT	840
TACGAATGGA	TGAAGTCGTT	ACGCCAACCG	CATCTACTAC	CATCGGGAGA	GAAGATTGGT	900
TTGCATACAA	AGCTGCCATG	CGGATTGCTT	TTTCCTTCTC	AGCTGACAAA	TGCCCCAAAT	960
TGATGAAGAG	AGCCTGGCTT	TGCTTAGTAA	AATCAAGAAC	TTCACGGGGA	TCATCTGCCA	1020
TGACAGGTTT	GCATCCCGA	GCCAAAATCC	CATTTGCCAG	CATCTCACAA	GAAATCTCAT	1080
TGGTCATACA	GTGAATGAGG	GAAC TAGAGC	CTATAGGAAA	AGGATTTGTC	AATGCCTGCA	1140
TCATTCTATC	CTTTCAGCAA	AGAAATATCC	TTGCACTTTT	TTAAAGAATT	CCTGCTTGAT	1200
TAAAAATCTA	AATGCAATAA	AGGAAATCGC	TGTACCAATC	AAGGTTGCTC	CGAAAAATCG	1260
AGGCGTGTAG	ATAAACCAAC	TAAGCTTAGC	AGCCGATCCT	GTAAAGAGCA	CCATAACAGG	1320
ATAGGAAACA	ATAGAACCAA	TAATACCTGT	TCCCACAATT	TCTCCCAAGG	CAGAAAAGTA	1380
AAATTTTCGA	CCGTACTTAT	AAAAGAGACC	TGCTAGAAGG	GCTCCAAAAG	TCGCTCCTGT	1440
GAGAGATAAA	GGAGCTTATC	GGAATACCCT	TGAGTCGTCA	TACGGATAAA	GGCTGTCACT	1500

GTAGCCATAG	CCAAGGCATA	AACAGGTCCC	ATCATGATTC	CCGCTAGAAT	ATTGACTACA	1560
CTGGACATCG	GTGCCATTCC	CTCAATCCGA	AAGATAGGTG	TAAGGACTAC	ATCAAGGGCA	1620
ATCATCATAG	ATAAAATGGT	CAATTTGTGA	ACTTGTAGTT	GGTGCTTTCT	CAAGTTTCTA	1680
TTCTTCTCCT	TTTTCTAAAG	ACTGTAAATC	GCTCTTCCAT	GTCTGGTGTT	GGTAAGCCAT	1740
CTCCCCAAAC	TTGGCTTCCA	TATGAACACT	GATGTGGAAG	GCATCTAGCA	TTTTTTGCTT	1800
ATCTGTCTCA	TCACTTTCTC	GATAGAG				1827

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 485 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

CTATTGCCAA	TCCATATAGC	CTATCAGGTG	GTCAATAACA	ACGTGTGGCC	ATCGCTCGTG	60
GCCTATCAAT	GAATCCAGAC	ATCATGCTCT	TCGATGAACC	AAATTCTGCC	CTTGACCCTG	120
AGATGGTTGG	AGAAGTAATT	AACGTTATGA	AGGAATTGGC	TGAGCAAGGC	ATGACCATGA	180
TTATCGTAAC	CCATGAGATG	GGATTTGCCC	GCCAGGTTGC	CAACCGCGTT	ATCTTTACTG	240
CAGATGGCGA	GTTTCCTGAA	GACGGAACAC	CTGACCAAAT	CTTTGATAAC	CCACAACACC	300
CTCGTCTGAA	AGAGTTCTTA	GATAAGGTCT	TAAACGTCTA	AACTCAAAC	GCAAGGATTT	360
CCTTGCAAGT	TTTCTACCTC	GTATTGGAAT	TTTTGATTTT	TCGGAAAATT	ATGTTAGAAT	420
TAAGTTTATG	AAATGAGGTT	TCCTCATACC	TAGCAAGACT	AGGAATAAAA	ATAGAAATTA	480
GGTAG						485

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1547 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

NTCTTGGGCN	CNGGGCGNNT	CCTTTGAGGA	CNACGGTATC	GATGACCTTG	ATCTCAAGTG	60
CAAGCAGTAT	CTGAATCTGC	AGCAGCACCT	GTCCGTGCAA	AAGTTCGTCC	AACATACAGT	120
ACAAACGCTT	CAAGTTATCC	AATTGGAGAA	TGTACATGGG	GAGTAAAAAC	ATTGGCACCT	180
TGGGCTGGAG	ACTACTGGGG	TAATGGAGCA	CAGTGGGCTA	CAAGTGCAGC	AGCAGCAGGT	240

TTCCGTACAG	GTTCAACACC	TCAAGTTGGA	GCAATTGCAT	GTTGGAATGA	TGGTGGATAT	300
GGTCACGTAG	CGGTTGTTAC	AGCTGTTGAA	TCAACAACAC	GTATCCAAGT	ATCAGAATCA	360
AATTATGCAG	GTAATCGTAC	AATTGGAAAT	CACCGTGGAT	GGTTCAATCC	AACAACAAC	420
TCTGAAGGTT	TTGTTACATA	TATTTATGCA	GATTAATTTA	CAGAGGGACT	CGAATAGAGC	480
CCTCTTTTCA	GGTTTTACCG	TGACAATCCC	TATTAATAAT	TATATCAAAA	TCGTGAAAA	540
ATTGGAAAAAG	TATGGTAGAA	TGAAAATTGT	CGTGTGAACG	ATAATACTCA	TTCTTGATGA	600
ATTGTGAAGC	AGTTGCCCTT	GGGTCGTTTT	GCGAGTTGAA	GTCAAGAAGA	GGAAAAAAC	660
AAAAAGGAGA	AATACTCATC	GAATTTCAAT	GAAACAAC	CTTGAGGCTG	GTGTACACTT	720
TGGTCACCAA	ACTCGTCGCT	GGAATCCTAA	GATGGCTAAG	TACATCTTTA	CTGAACGTAA	780
CGGAATCCAC	GTTATCGACT	TGCAACAAAC	TGTAAAAATAC	GCTGACCAAG	CATACGACTT	840
CATGCGTGAT	GCAGCAGCTA	ACGATGCAGT	TGTATTGTTC	GTTGGTACTA	AGAAACAAGC	900
AGCTGATGCA	GTTGCTGAAG	AAGCAGTACG	TTCAGGTCAA	TACTTCATCA	ACCACCGTTG	960
GTTGGGTGGA	ACTCTTACAA	ACTGGGGAAC	AATCCAAAAA	CGTATCGCTC	GTTTGAAAGA	1020
AATTAAACGT	ATGGAAGAAG	ATGGAAC	CGAAGTTCTT	CCTAAGAAAG	AAGTTGCACT	1080
TCTTAACAAA	CAACGTGCGC	GTCTTGAAAA	ATTCTTGGGC	GGTATCGAAG	ATATGCCTCG	1140
TATCCCAGAT	GTGATGTACG	TAGTTGACCC	ACATAAAGAG	CAAATCGCTG	TAAAGAAGC	1200
TAAAAAATTG	GGAATCCCAG	TTGTAGCGAT	GGTTGACACC	AATACTGATC	CAGATGATAT	1260
CGATGTAATC	ATCCCAGCTA	ACGATGACGC	TATCCGTGCT	GTTAAATTGA	TCACAGCTAA	1320
ATTGGCTGAC	GCTATTATCG	AAGGACGTCA	AGGTGAGGAT	GCAGTAGCAG	TTGAAGCAGA	1380
ATTTGCAGCT	CCAGAACTC	AAGCAGATTC	AATTGAAGAA	ATCGTTGAAG	TTGTAGAAGG	1440
TGACAACGCT	TAATTTATAC	AAATAGTAAT	TACCTAGGAG	GGCGGGGCTT	AGCCCGGCTC	1500
TCCTATTTTC	AAAAAATATA	GGAGAATTAA	AATGGCAGAA	ATTACAG		1547

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 740 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

CTATAAAAAA	AAGGGTAACC	AGTATGGAGG	ATGAATGTCT	GGAAC	TATCT	GAGAATCTCG	60
GATTTTGGAA	ATCAGACCGA	TCATCATGAG	ATAAGGAAGG	AAAGCACTTG	TAAAAAGCAC		120
TGTAACACG	CCAGTCCCCT	GTCCCAAGAG	GGTGAGGTGG	TAGCGTAAAA	CCATGCGGAA		180
AAATCCCTTT	TTAGTGGTTG	AAATTCCTCTC	CTTGCTGCGA	CGTTCTTTTT	TGACCTTCTC		240
CTCACTATTA	AGCAGGATCA	CGTCATAAAA	ACGAGGAAGG	ACCTTCTTTT	TGGTCAGATA		300
AAGCAGGAAG	AGAGTTAGTC	CTATCCAAGC	GAGCAGACCC	AATATGGCTT	CTATTGAAAA		360
AGGCTCCACT	GCTATTTTGT	AAAAGATATG	AAGAGGATAA	AGGAGAAATG	GAATGTCTCT		420
AACTTTGTCA	ACAATACTTC	CAAAAGTCGA	CTGAAGAAAG	AAGATAAATA	TAAAGGTAT		480

GAGAACTCCT	ATCCCAATCA	TCACATTCGA	AAAAATAGAC	TGATACTTTC	TGAAGACCCT	540
AGTCTGAGCC	AAGAAATGTA	CTGCCACTAC	CGTCACTAAA	GTAACAGAGA	CAAATAATAA	600
GGTCAAGGAC	AGTAGCATCA	AAGGCAAACC	CAGCCAAAGA	GAAGGAGCTA	GACTAATATA	660
GAGGGCTAGA	AAATAAGCTA	GGATTGGTAC	AATTCCAGTT	AGAGCTGGCA	AGAGGACAGA	720
CAGTCCTTTA	GCAATTCGAT					740

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2219 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

ATCGAATTTCG	TTTTGCAAGT	GGCGAAATGC	GAACCACGTT	TGTGTCTTTA	TAAGTTTCCA	60
CGTCTTCTTT	GTGGACACGA	CCGTTTGCAC	CTGAGCCAGA	AACGTCGTAG	AGGTTTATCC	120
CTAAATCATC	CGCTAACTTT	CTAGCTGCAG	GAGTCGCTCT	TAGCTTGTC	TCAGCCATGA	180
CCTCTCCAAT	TCTATTTATG	ATACAAAGGG	CGTCAAAAGC	GACTGAAAAA	TAGGAAATCG	240
ACGATGGCTT	CGATGAAGCC	AAGGAGATTT	ATCTTTTTTT	CCAAGCTTTT	AGCCCGTGCT	300
CTAATCTAAG	ATATTAAGGA	CGAAGAGCTC	TGCACCTAAA	AGATACAAAG	TTCTCGTCAG	360
CTTTGTTTTA	TTTACATAAC	TTATCTTATG	TAACCTCTAT	CTTTGTTATA	AGTTTTTCGG	420
ATTGCATCTT	TGATACTTTC	AACTGTTGGA	ATCATTGCAC	ATTTTATAGG	TTTGCGCATA	480
AGGCATCGGC	ACATCTTCTC	CTGCACAACG	GCGGATTGGT	GCATCTAGAT	AGTCAAATGC	540
TTCTGATTCT	GAAATAATAG	CTGAAATTTT	ACCGATATAG	CCACTTGTTT	TGTGGGCATC	600
GTTGACCAGA	ACAACCTTAC	CAGTCTTCTT	CACTGAGTTT	ATGATGATAT	CCTTATCAAG	660
CGGAACAAGG	GTACGTGGGT	CAACAATTTT	AACTGAAATT	CCTTCTTCAG	CTAATTCTTC	720
AGCAGCTTGA	ACCACACGGC	GAAGCATTTT	TCCATAAGTG	ACAACCTGTT	CATCCGTTCC	780
TTGGCGTTTG	ATTTACACAA	CCCCAAGTGG	AATTGTGTAG	TCTGGATCAA	CTGGCACTTC	840
CCCTTTTTTG	TTAAATTTCT	ACTTGTAATC	AAGTATAATA	ACTGGGTTGT	TATCACGGAT	900
AGAAGACTTA	AGCAGGCCTT	TCATGTCCGC	AGGTGTTCCA	GGTGCCACAA	CCTTAAGCCC	960
TGGAATGTGA	GTAAACCAAG	ACTCTAGAGA	TTGTGAGTGC	TGGGCGGCAG	AGCCAACCTC	1020
GTTACCAGCT	GCACAACGAA	CAGTCATTGG	AACCTGACCT	TTACCACCAA	ACATGTAACG	1080
TGTTTTAGCA	GCTTGGTTGA	CGATATTGTC	CATGGCAATA	ACAGAGAAGT	CCATGAAGGT	1140
CATATCGACG	ATTGGACGAA	GTCCTGTCAT	GGCTGCTCCT	GCTGCAGCTC	CAGAGATGGC	1200
AGCTTCAGAA	ATCGGACAGT	CACGGACACG	TTCTGGACCA	AATTCTTCAA	GCATTCCAAC	1260
AGAAGTACCG	AAGTCTCCTC	CGAAGACACC	GACGTCTTCT	CCCATCAAGA	ACACATTTTC	1320
ATCGCGAACG	CATTTCTCTA	GACATAGCAA	GGATAATGGT	GTCACGGAAG	GACATTGTTT	1380
TTGTTTCCAT	TTTATCTCTT	TCTCCTTAGT	CTGCGTAAAT	ATCTTCAAAG	GCTGATTCAA	1440
GCGGTGGGAA	TGGGCTTTCC	TCTGCAAATT	TAACAGAAAG	TTCTACTGCT	TCCTTTACTT	1500
GCGCTTGGAT	TTCTTCCAAT	TCTTCGGCAC	TTGCAATGTT	ATTTTCAATA	AGGTAATTGC	1560

GGAGGTTTTTC	GATTGGATCT	TTTTGTTTTCC	ACAATTCCAC	TTCTTCACGC	GTACGATATT	1620
TACCAGGGGTC	AGATGATGAG	TGACCGAGCC	AGCGATAAGT	TACACTTTCA	ATCAAGACTG	1680
GACCATTGCC	ACTGCGAACA	TGGTCTATAG	CTTTCTGAAA	TCCTTCATAG	ACATCGATGA	1740
CATTGTTACC	GTCTTCGATG	AACATTCCAG	GAATTCCATA	AGCGGCGCTA	CGTTGATGGA	1800
TATGTTCTAT	ATTGGTCATT	TTCTTGATAT	CCGCAGAGAT	ACCGTAACCG	TTGTTAATGC	1860
AATAGAAAAT	GACTGGCAGG	TTCCAGATAG	AAGCCATGTT	CACTGCTTCG	TGGAAAACAC	1920
CTTCATTGGT	CGCACCATCT	CCAAAGAAGC	AGACAACGAT	TTTACCGGTA	TTTTGCATTT	1980
GCTGACTGAG	GGCTGCACCG	ACAGCGATCC	CCATACCACC	ACCTACGATA	CCATTGGCAC	2040
CAAGGTTCCC	AGCATCAAGG	TCAGCGATAT	GCATAGATCC	ACCTTTCCCT	TTACAGGTTC	2100
CAGTGATTTT	ACCAAGGATT	TCAGCCATCA	TTCCGTTGAA	GTCAATCCCT	TTAGCAATAG	2160
CTTGCCCGTG	TCCACGGTGG	TTTGAGGTAA	TCAGATCATC	TGGATTGAGA	GCTACATAG	2219

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1078 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

CTAACCCCTNG	ACGGGGCCGC	TATCATCAGT	CAAACAGCTA	AAAATCTTGT	CTGCAAAAGT	60
CTCGATTAAAC	TGAGCTTTTA	CAAAAGCCGT	ATTTCTTGGA	ATAACTTGGA	GATTGATCAT	120
CTTATCCATC	AATTCAGCCG	ATTCGATATT	GTCTTCAGCC	AGTTGCAGAC	TTTTTACGAT	180
TGATTTTGGC	AATTCGTAGA	CATAGGTGTT	GTCTCTCAAA	GGAATTTTGA	CAATACCTAA	240
CTCTTTGATA	TCTCGGGATA	CCGTCGCCTG	AGTGGCAGTG	ATACCTGCTT	CTTTCAAATG	300
TTCTACAATT	TCTTCTTGCG	TGCCGATTTG	ATAATCTGTC	ACCAATCTTC	TAATTTTTTC	360
AAGTCTCTCT	TTTTTTATTCA	TTTTTTAAAT	GACTATGCGC	CCTCTCTACT	GCTTCTTTAA	420
TCTCAGCAAG	AATCTGATTG	CTTGCTGACT	TTTCTTTTTT	CAAATACACT	AAAAATTCAA	480
TATTTCCATG	TCCACCTTGG	ATGGGAGAAA	AGTCCAAGCC	AAGGACTGAA	AAACCTGCCT	540
CTACTGCCAT	AGCTGTTACA	GATTCAAGGA	CATTCTGATG	AATCTTAGCA	TCTCGAATAA	600
TTCCATTTTT	CCCAATCTGC	TCACGTCCTG	CCTCAAACCTG	AGGTTTGACA	AGTGCTACCA	660
CCTGACCTTG	ATCAGCCAAG	ACACGGTGCA	AGGCTGGCAA	AATCAGACTA	AGGGAAATGA	720
AACTCACATC	AATACTGGCA	AAGCTCGGCT	CCTGCTCGAA	ATCAGTCTTT	TCAGCATAGC	780
GGAAATTGAA	CTGCTCCATG	CTGACAACCTC	GTGGGTCTTG	GCGTAATTTT	CAAGCCAACCT	840
GATTGGTACC	AACATCGACT	GCAAAGACCA	ACTTGGCACT	ATTCTGTAGC	ATGACATCGG	900
TAAAACCTCC	AGTAGAGGCC	CCGATATCAA	TCGTAGTCGC	GCCATCCACC	GACAAATCAA	960
AGACCTGCAA	GGCCCTTTTC	CAGTTTCAAA	CCACCACGGC	TGACATACTT	GAGTTTCTCC	1020
CCCTTGAGTT	TTAATTCGGT	GTCATCTGGA	ATTTCTCTCC	TGGCTTGTC	AACCGTTC	1078

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 928 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

ACTTTCCTGA	CCTCTGTTTC	CAAATAATCT	TCCAAATGGA	CAGAGATCTA	CCGTTGTTTG	60
CATCGATAGC	TGAGGTCTTT	TTTAGAAAAT	ACCATCACTT	TTAGAAAATA	TAAACACATT	120
TTTCGGATAA	GATTAAGGTT	AAAAGCAGCT	CGTTTATCCA	GGGTCTGATG	ATGGTCTTCA	180
CGATAAACCA	CATCCAATAA	CCAATGCATA	CTTCTGCTG	ACCAATGACC	TCGAACACTA	240
TGGCAAAAGG	TCATCAACAT	CAAGCTTAAA	GTTAAAGATA	AAATAGCGAA	CGTCTTGACT	300
TGTAATACCA	TCTCTATCAA	TAGTATTACG	AGTCATTCCA	ATTCCACGCA	ATTTATGCCA	360
TTTGGGATGG	TTTTGACACA	ACCACTTAAC	ATCAGAAGAC	ACCCAGTATT	CTCGAACTTC	420
AATCTATCCT	CTTCTATAT	TCTAACTGAA	AGGACAATTC	AATGATTTCAT	TTAATAATGA	480
TTAGCGCCAT	TGCTCTAGCC	ATTGGAATTG	GTTACCGCAC	CAAAATCAAT	ATTGGCCTGC	540
TGGCTATTGC	TTTTTCTTAC	CTCATCGCAA	CCACTCTCAT	GGGATTAAGT	CCCAAAGAAC	600
TTCTTCATTT	TTGGCCAACC	TCACTCTTTT	TTACCATTTT	TAGCGTCTCT	CTCTTTTATA	660
ACGTTGCAAC	AACTAACGGT	ACTCTTGATG	TTTTGGCTCA	ACACATTCTC	TACCGCACAC	720
GCACCCACCC	TAACGCCCTC	TACATGATTT	TATACCTGAT	GGCAACCCTT	TTGTCTGCTT	780
TAGGTGCTGG	ATTTTTCACT	ACTATGGCCG	TTTGCTGTCC	TCTAGCGATT	ACCCTCTGTC	840
AAAAAGCGGA	CAAACACCCT	TTGATTGGAG	TCAAAGCGTC	AATGGGAACT	TCAGGAAGGG	900
TAATTTGATA	ACCAAAGGAA	TAAAATTT				928

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 847 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

AAAAACGCAC	CATATCAAAA	ACTAAAAAGT	TTGATATCAT	GCGTCATGTC	TTAAACTAAT	60
TGACTATACT	TTCTATTCAA	ATGAGCTTTT	AACCAATTGA	TTGAGCCAAT	CCACTCTTAA	120
AACCAAAGGA	GCAATTTCTC	GGCTTAGCTG	ACTCTTCTCG	GAATCTGAAC	CATGTACAAC	180
ATTTTGGATA	ATCTCATTTT	CTCCAGCAGC	TTTTGCAAAA	TCACCTCGAA	TAGTGCTTGG	240
TAAAGCTTCT	TCTGGACGAG	TTGCACCCAT	CATGGTCCGC	CAAGTTTCGA	TTACTTTGGG	300

ACCAGAAATG	ACACCCACAA	GAAGTGGACC	TGAAGTCATG	AATTCACGAA	TCGGTGGGTA	360
AAAACCTCTGA	CCAACCAAGT	CCTGATAGTG	CTGGTCAATC	AACCTCTCTG	AAAACCTGTG	420
AACGAAACTC	CAATTTTTCG	ATTGTAAATC	CACGTTGTTC	GATGCGCTTT	AACACTTCAC	480
CCACTAGCCC	TCTTTTACAA	CCATCTGGTT	TGATGATAAA	GAATGTTTGT	TCCATACCCG	540
TCTCCTTTGT	CAGCTTCTTT	CTTTTATTTT	ACCACATCTC	GTGGAAAAAT	GGAGAAAGTT	600
TTCAGAAGAG	AGAATGAGAG	AACCCTCGGG	TTCTCTCATT	CTCTCTTATT	CTACTGTTTC	660
TTCCACAGTG	TCAACGGCAG	TATCCACAAC	TACTTCTGTT	GTTTCTTCAT	TTCTTCTTTC	720
CTCTACTGGA	GGATTAAGGT	ATTCTTCTTC	GTTGACAGCA	TGTGGTTCAA	GGTTACGGTA	780
ACGGGCCATA	CCAGTACCAG	CTGGGATGAT	CTTACCGATG	AATAACATTT	TCCTTTAAAT	840
TCCAAGG						847

(2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 578 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

ACAACCTAAC	TACCGNCTAA	TTCAGCGCGA	ACTTCTGCAG	TAGCTGCTTC	AACAACCTCA	60
CGACGTGAAA	GGATGAAGCG	GTTTTCTTTA	GCGTTAACTT	CTTTGATTTT	AGTATCAAAAT	120
TCTTGACCTA	CAAAACGCTC	AGCGTTACGT	ACGAAACGAG	TATCCAACAT	TGAAGCTGGG	180
ATAAATCCAC	GAACACCTTC	AAATTCTACT	GAAAGTCCAC	CTTTAACGGC	ACGCGTTCCT	240
TTAACAGTAA	CAACTTCTTC	TTCGCGACCA	ACAAGTTTGT	CCCATGCTTT	GCGAGCTTCA	300
AGGCGTTTTT	TAGATGACAA	GGTATGTAAC	TGTATCAGTA	TCTTTACCAA	CTACTTGACG	360
AAGTACAAGA	ACATCCAATA	CTTCTCCTAC	TTTAACAAAG	TCATTGATAT	CTGCATCACG	420
ATCGTTTGTC	AATTCGCGAA	GAGTCAAGAC	ACCTTCAAC	ACCAGTCCC	AGAAGAATGC	480
AACGTTAGCT	TGAGTCGCAT	CAACTGTCAA	TACTTCAGCA	CTAACACATC	ACCAGTCTCA	540
ACTTGACTNA	CGCTATTGAG	CANATCTTCA	AATTCGAT			578

(2) INFORMATION FOR SEQ ID NO:26:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 888 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

GTAGTTATAG	TAGGGGTCGG	ATTGAAATGC	CACNGCGCTT	CTTGGAGTTT	CTGATACCGT	60
TTAAAATAGC	GTTGGGCATT	CTGGTTGGGA	GTCAGAGCCT	TATCAAGCGC	AATCATGATA	120
GGTTGGTTGG	TATAGTAGTT	GTCTAGGATA	ACCTGGTTCT	TGGTCGTTAG	GCACCTGGTG	180
GAGGAAGGTT	GTCAGCAATT	CTCCTTTTTT	ACGAAATTCT	TCAGCGTTGT	CTGTCGCCAG	240
TAACTATTTT	TCCTGTTTTT	TGAGTTTGTG	TCGGTTTTTC	TGAAGTTCAT	TTTCAACACG	300
ACGAATCAGT	TCCTGGCCT	GCTGTTTGAC	GCGGTCGCGC	TCAGCCTTAT	CCTTATAGTA	360
GGTGTCCAAC	AAATCAGAAA	GATTTGCAAA	AGGCTCTCCC	ACCTGATTTG	CAAAAGGAAC	420
TGGACTGAAG	GAAGTCTCAG	TCAAGCATGG	CTTGGTTTCC	TGATTGAAAA	AATTTTCGGAA	480
AGCGGAAAGT	TTTTCACTAA	CCAGTATCCT	TTCCAATTCA	TTTGCCGTAT	CGCGTCCCAG	540
ACCTTGAAAG	AGGCTTTGAA	GATTTTTTGC	TGTTAGTTCT	TGGGTTTGCA	GGATTTCAAA	600
GAGCTTTTCA	TCCTTGATAG	TAAAAGGATT	GAGAGATTCT	GTACTTGCGC	GAGCGATATA	660
GGTCGATCCT	GGAAGTAAGG	TGCGGTAGCT	ATTTTGTGAA	AAGCCGACGT	GTTTGATAAC	720
TTCGAGGATT	TTATGACTGC	TTTTATCCGA	CCAGTTAGAA	TATTACTGTG	TTTCCCCATA	780
ATTTCGATAA	TCAAGGTAGC	CTGGATATGG	TCTCCAATCT	CGTTTTTTAT	GGAAACTGTA	840
ATTTCCACAA	TACGGTCATT	TTCCACTTGC	TCAATCGACT	CAATCAGG		888

(2) INFORMATION FOR SEQ ID NO:27:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 513 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

ATCGAATTTT	GTTCTTTCAT	AGAGAGCTAC	CTGAGTTCTA	TTCAAGCTCA	GGTAGTACTT	60
TCTTATAAAC	TAGACAAACT	AACTGTCATT	CTACCATCAG	ATTACAAGAC	ATCATCGTCA	120
CTCACCTTGG	AATTCAATGT	CGTACCCCAA	TGGGTAATTT	TACGGTGGGG	TTGAGCTAAA	180
ATTGGTCTGT	TTTCATAGAT	TGTTTGCCAT	CTATTCCATA	GTAGGCCCGT	CTTTTCTCTCA	240
ATCTTAACTC	GCAGATTTCT	CATATTTTCT	TTGATTGGGA	GGTTGAGGAC	AAAACCTGCA	300
GTCTGGTTGC	GACCGTTTCC	TTCCCAAGAA	TGACTACGAA	CAACTTGGTT	TCCATCTTTA	360
TCTACTGGAA	CTTCTTCCCA	AGTTATGGAG	TAGCGGGCAA	TGTAAGCTCC	ACTGTGTTGA	420
ATTATCAATG	TTTTATCTTT	CACAGGGAGT	CTGACTGATT	GGTTGAACTG	GCTTAGAAAC	480
TTGTGTCGCC	GTTTCAGCAT	TCGTAGCTAT	AAA			513

(2) INFORMATION FOR SEQ ID NO:28:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 214 base pairs
- (B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

ATCGAATTCT AACATGTGCT TCTCCTTCTA TTGTTTCCTAT CTTTAAAATC TACTCCTTCA	60
TGCTCCAAGA GCCAAGCTTT CTTTTCCACT CCTGCAGCAT AACCTGTCAG ACGCTTGCCCT	120
GCTCCCAACA CACGATGACA AGGTACTAGG ATAGACCAAG GATTGCGTCC CACTGCTCCA	180
CCAATTGCTT GAGCAGAAGC CACTTGCAGG TCTT	214

(2) INFORMATION FOR SEQ ID NO:29:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1084 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

CTCCAGCAAT GGATCCAAGT ATGATGGGCG GGATGATGTA AGCTTTCTAT AGAAAACACC	60
TTATAAAAAA CACGAAAGGA GGGAATGACT AACCTTCTTT TTTATAATAT TCACTTCTAA	120
GATTGATGGT GAGCTCTCCT AACTTATATG ATAAAAATAAG ACTAGAGGAA AGGAGAAGAA	180
CATGATCGAT GTACAAGAAA TTCTGTGCAA GATGACCCCC AATCAGAAGA TTAATTATGA	240
CCGTGTCATG CAGAAAATGG TACAAGCATG GGAAAAAAAT GAGTAGCGGC CAACCATTCT	300
CGTGTCATGTT TGCTGTGCCC CTTGTAGTAC CTATACACTA GAATATTTGA CCAAGTATGC	360
AGATGTGACC ATCTATTTTG CCAATTCTAA TATCCATCCC AAGGCAGAAT ACCATAAGCG	420
GGTCTATGTC ACCAAGAAAT TTGTTAGTGA TTTTAATGAG CAGACAGGAA ATACGGTTCA	480
GTACCTAGAA GCTCCCTACG AACCCAATTA ATACCGAAAA CTAGTTAGGG GGCTAGAGGA	540
GGAGCCCGAA GGTGGCGACC GTTGCAAGGT TTGTTTTGAC TACCGACTGG ATAAAACAGC	600
GCAAGTGGCT ATGGACTTGG GCTTTGACTA CTTTGGTTCA GCCTTGACCA TCAGTCCTCA	660
TAAGAATTCT CAACTATCA ATAGCATCGG AATCGATGTG CAAAAAATTT ACACGCCCCA	720
CTATCTTCCC AACGATTTCA AGAAAAATCA AGGCTACAAA CGTTCAGTAG AGATGCGTGA	780
GGAGTATGAT ATCTATCGTC AATGTTATTG TGGCTGCGTC TATGCAGCCC AAGCCCAGAA	840
TATTGACCTG GTTTAAGTTG AGTAGGACGC CACAGCATGC TTGCTGGATA AGGATGTTGA	900
GAAAGACTAT TCTCATATCA CATTTATAGT AGATTGAAAC TAGAATAGTA CACCTTTACT	960
TCTCAAACAT TGTTAGAAAT CGATTCGGCT GTCCCTTATTT CATTTTAATA TACTGGTACG	1020
AAATTAGATA TATCAATGAT AACTTGCCTC AAGGTAGGTT TTTTGATAGT AGAAAAGCGA	1080
TAGA	1084

(2) INFORMATION FOR SEQ ID NO:30:

144

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1124 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

```

ATCGAATTCA TTGACTGCCT GAAAAGACTT CAACTCGTCT GCCTGATAAC CGAAAAGACTT      60
GGTTACTTTG ATACCTGATA CGGACTCCTG TACCTTGTTA TTGAGTTCAG AAAAAGCAGC      120
TTGGGATTCG CCAAAGGCCT TATGAGTCTT TCTCCCTAGG CGACTAGTCG TATAGGCCAT      180
GAAAGGTAGG GGGAGAATGG CAACAAGAGT CATCTGCCAT GAGATGCTAA AGAGCATGGT      240
CAACAAAGTC ACCAGAGCCG TGATAGAGGC ATCCACCGCA GACATGACAC CGCCACCTGC      300
TAAACGAGTC AAGGAATTGA TATCATTGGT TGCCTGTGCC ATCAGATCAC CCGTCCGATA      360
GGTTTGATAA AAGGCTGACG ACATTTTGTG GAAATGCTTA AACAAGCGAG ACCGCATGAT      420
CTGTCCCAAG CAATAAGAGG TCCCAAGGAT ATACATACGC CACACATAGC GCAAATAGTA      480
CATACCAAAG GCTGCAAGTA GCAAGTAAAA TAGGCTAAGA AGGAGGTCCT GCTGGGTAA      540
TTGCCCCGAT GTGATGGCAT CAATAACCCG CCCCATTAACC ATAGGAGGAA TGAGATTGAG      600
GACGGAAACC AAGACCAGGG CCACAATCCC GACTAGATAA CGGCGTTTTT CTAACCTGAA      660
AAACCACCAA AATTTTGTAA TAATGGACAT AAAATCCCTT TCTGGATTGC AAATAGAAAC      720
CTGAGGCCAA TACTCAATGG AAAATCAAAG AGCAAAC TAGGAACTAGC GCAGGCTGCT      780
CAAAGCACTG CTTTGAGGTT GTAGATAGAA CTGACGAAGT CAGTAACCTA CACACGGCAA      840
GGCGACGTTG ACGCCGTTTG AAGAAATTTT CGAAGAATAC AAGACCCAG GTTTTTCTTA      900
TTTATAAGTT ACCACTGTAA CAGCACCTTT GTCATATTCA GCAATAAAGA TATTGGCTAC      960
ATTGTCATGC CCTTGTTTAC TGAGGTTATC AAGCAACCAC TCCTCGCTAC GAACAATCGA     1020
TCCCAAGACA TCTACTTGAA TCACACCGTC AGTCACAACT GGATACTTAG GATTTTCATC     1080
TCCCATTGTC ACAACGATGA GTTGCCCATT TTGCTCTTGC ACAG                          1124

```

(2) INFORMATION FOR SEQ ID NO:31:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1242 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

```

TTACCTTCAT TGCAGCCATT ATTGGTTCTT GTGTCAGCCA GATTTTAAGT ATTCTTTATA      60

```

AGACACCTGC	TGTGGTCTTT	ATCTTGGCCA	TTTTGGCACC	GCTGGTTCCA	GGTTATCTCT	120
CCTACCGAAC	AACTGCCTTT	TTTGTGACAG	GGGACTATAA	TAAAGCACTG	GCAAGTGCGA	180
CCTTGGTTGT	CATGTTGGCT	TTGGTAATCT	CTATTGGAAT	GGCTAGCGGA	ACAGTGATTC	240
TCAGACTGTA	TCATTATATA	AAAACACATC	GAGTATCGTA	GACTTTACAG	AAATAAAAAGA	300
ATTTTCTGAA	AAATGAGATA	AATAAATTAA	CAACGCTTTC	TATATGTGCG	AGAATACCGC	360
ACTTATGAAG	AAATTGCGGC	TGATTTTGGT	ATCCACGAAA	GCAACTTAAT	CCGTGCGAGC	420
CAATGGGTTG	AAGTAAC'TCT	TGTTCAAAGT	GGTGT'TACGA	TTTCAAAAAC	TCATCTTAGT	480
GCTGAGAATA	CGGTGATTGT	GGATGCAACA	GAGGTAAAAA	TCAATCGCCC	TAAAAAACAA	540
TTAGCGAATG	ATTCTGGTAA	AAAGAAATTT	CACGCTATGA	AGGCTCAGGC	GATTGTCACA	600
AGTCAAGGGA	GAATTGTTTC	TTTGGATATC	GCTGTGAACT	ATTGTCATGA	TATGAAGTTG	660
TTCAAAATGA	GTCGCAGAAA	TATCGGACAA	GCTGGAAAAA	TCTTGGCTGA	TAGTGGTTAT	720
CAAGGGCCCA	TGAAGATATA	TCCTCAAGCA	CAAAC'TCCAC	GTAAATCCAG	CAAAC'TCAAG	780
CCGCTAATAG	CTGAAGATAA	AGCTTATAAC	CATGCGCTAT	CCAAGGAGAG	AAGCAAGGTT	840
GAGAACATCT	TTGCCAAAGT	AAAAACGTTT	AAAATGTTTT	CAACAACCTA	TCGAAATCAT	900
CGTAAACGCT	TCGGATTACG	AATGAATTTG	ATTGCTGGCA	TTATCAATTA	TGAACTAGGA	960
TTCTAGTTTT	GCAGGAAGTC	TATTATTTTC	CTTATTGTCT	GTAAGTCTAC	TGACCTTGTT	1020
GTTTATCCCA	GTCATGGTTT	CTAGTTCGGG	CTCAGAGTTT	CAAAGTGGAT	GGCAAGAGCA	1080
TCAATTGATT	GCTGAGAAGG	TTAGTAAAC	ACTTGACAAG	ACATTTGATA	AGGATGTCAG	1140
AAAAAT'TCCG	ACCAGTCAGT	TTTATCAAAA	ATTTGTAGAT	GAGATGGGAA	GGATTTACTC	1200
AGGAAATTTG	ATCCTCCCAG	GAGCTGATAA	CTGTGAATGG	AG		1242

(2) INFORMATION FOR SEQ ID NO:32:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1575 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

GTGATGGGGC	CTCAGGGAAA	TGGTTTTGAC	TTGTCTGACC	TTGATGAGCA	GAATCAGGTT	60
CTCCTTGTTG	GTGGTGGGAT	TGGTGT'TCCA	CCCTTGCTTG	AGGTGGCCAA	GGAATTGCAT	120
GAACGTGGAG	TGAAAGTAGT	GACAGTCCTC	GGTTTTGCTA	ATAAGGATGC	TGTTATTTTG	180
AAAACCGAAT	TGGCTCAGTA	TGGTCAGGTC	TTTGTAACGA	CAGATGATGG	TTCTTATGGC	240
ATCAAGGGAA	ATGTTCCGTT	GTTATCAATG	ATTTAGATAG	TCAGTTTGAT	GCTGTTTACT	300
CGTGTGGGGC	TCCAGGAATG	ATGAAGTATA	TCAATCAAAC	CTTTGATGAT	CACCCAAGAG	360
CCTATTTATC	TCTGGAATCT	CGTATGGCTT	GTGGGATGGG	AGCTTGCTAT	GCCTGTGTTT	420
TAAAAGTACC	AGAAAGCGAG	ACGGTCAGCC	AACGCGTCTG	TGAAGATGGT	CCTGTTTTTCC	480
GCACAGGAAC	AGTTGTATTA	TAAGGAGAAA	ATTATGACTA	CAAATCGATT	ACAAGTGTCT	540
CTACCTGGTT	TGGATTTGAA	AAATCCGATT	ATTCCAGCAT	CAGGCTGTTT	TGGCTTTGGA	600
CAAGAGTATG	CCAAGTACTA	TGATTTAGAC	CTTTTAGGTT	CTATTATGAT	CAAGGCGACA	660

ACCCTTGAAC	CACGTTTTGG	GAATCCAAC	CCAAGAGTGG	CAGAGACGCC	TGCTGGTATG	720
CTCAATGCAA	TTGGCTTGCA	AAATCCTGGT	TTAGAGGTTG	TTTTGGCTGA	AAAGCTACCT	780
TGGCTGGAAA	GAGAATATCC	AAATCTTCCT	ATTATTGCCA	ATGTAGCTGG	TTTTTCAAAA	840
CAAGAGTATG	CAGCTGTTTC	TCATGGGATT	TCCAAGGCAA	CTAATATAAA	AGCTATCGAG	900
CTCAATATTT	CTTGTCCCAA	TGTTGACCAC	TGTAATCATG	GACTTTTGAT	TGGTCAAGAT	960
CCAGATTTGG	CTTATGATGT	GGTGAAAGCA	GCTGTGGAAG	CCTCAGAAAGT	GCCAGTTTAT	1020
GTCAAATTAA	CCCCGAGTGT	GACCGATATC	GTTACTGTCTG	CAAAAGCTGC	AGAAGATGCG	1080
GGAGCAAGTG	GCTTGACTAT	GATCATACTC	TGGTGGGATG	CGCTTTGACC	TCAAAACCAG	1140
AAAACCAATC	TTGGCCAATG	GAACAGGTGG	AATGTCAGGT	CCAGCAGTTT	TCCAGTAGCC	1200
CTCAAACCTCA	TCCGCCAAGT	AGCCCAAACA	ACAGACCTGC	CTATCATTTG	AATGGGGGGA	1260
GTGGATTCTGG	CTGAAGCTGC	CCTAGAAATG	TATCTGGCTG	GGGCATCTGC	TATCGGAGTT	1320
GGAACAGCTA	ACTTTACCAA	TCCTTATGCC	TGCCCTGACA	TCATCGAAAA	TTTACCAAAA	1380
GTCATGGATA	AATACGGTAT	TAGCAGTCTG	GAAGAACTCC	GTCAGGAAGT	AAAAGAGTCT	1440
CTGAGGTAAA	CTGCAATCAA	TCTGTTCTTG	ATTTTTTATT	AGTTTGTAAT	ATGAATTTAG	1500
GAGAATTTTG	GTACAATAAA	ATAAATAAGA	ACAGAGGAAG	AAGGTTAATG	AAGAAAGTAA	1560
GATTTATTTT	TTTAG					1575

(2) INFORMATION FOR SEQ ID NO:33:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 776 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

CTAAGATATC	AGAATAACAA	CGAAATCGAA	GCATTAAAAA	CAAATATTAC	TTCTAAGAAT	60
AGCGAGATTG	ATAGTCAACA	AAGCAATATT	AAGGATATGA	CCGTACCTAT	AATGATCCAA	120
CTTCTCAGGC	TTATAATATT	TATGCTCAAT	TAATTAGTGA	GTTAGGTAAT	GCTCGTTCAA	180
ACAACAATAA	AAGTATTACA	GAGCTTGAGG	CTAATCTTGG	AGTGGCAACA	GGTCAAGATA	240
AAGCTCATAG	TATATTAGCG	TCAAAATGAAG	GTACTCTGCA	TTATCTGGTA	CCTTTGAAAC	300
AAGGAATGTC	TATTCAGCAG	GGGCAAACGA	TAGCAGAAGT	TTCAGGGAAA	GAAAAAGGTT	360
ACTATGTAGA	GGCTTTTGTA	CTTGCGAGTG	ATATTTCTCG	TGTTTCAAAA	GGAGCAAAAAG	420
TTGATGTTGC	TATTACTGGT	GTGAATAGTC	AAAAATATGG	AACACTAAAG	GGACAAGTCA	480
GACAGATTGA	TTCAGGAACA	ATTTCCCAAG	AAACGAAAGA	GGGGAATATT	AGCCTCTATA	540
AAGTCATGAT	AGAATTAGAA	ACCTTAACTC	TAAAACATGG	AAGCGAGACG	GTCATACTCC	600
AAAAGGATAT	GCCAGTTGAA	GTGCGGATTG	TCTATGATAA	AGAAACCTAT	CTTGATTGGA	660
TTTGTAGAAAT	GTAAAGTTTC	AAGCAATAAT	TGGTTTTTAA	CCTTAGGTAA	CCTATAAAAA	720
CAAATAAGGT	AGAGAAAGGA	TATTTTATCT	AAGTTAGCTC	ACATTACTGC	CATTCC	776

(2) INFORMATION FOR SEQ ID NO:34:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1487 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

CTGGCCTTTC	TCCACCAAAA	TTGTTCTTTC	AGGGAAGGAA	GTCAGAACAC	TAGCCGTTGC	60
ATCTTCCTTT	TGCTTTTCAA	TCGTAATTCC	AGATAATTTT	TCCCATTCTT	TTTGGTGACC	120
CCGGGAGGCA	GGATTGAATG	GCTTGAGGGA	AATGACAAAC	TTGTCCTAGC	AAGAATGGTC	180
AAGGCACCTC	CGTCTACAAT	CAAAATCTGA	TTTGGGCTTA	AATTAACAAA	GACCTGTTTT	240
ACTAGATTTT	CTCCAGAAGC	ATCGTCTCGT	AAACCAGGCC	CCAGCAAGAT	AACTTCTGCC	300
TTCTCCAATT	GCTCTTTTAA	CAATTGCTGG	TCTTGAAGAG	AAAAGGCCAT	AGGCTCAGGT	360
AAATGGCTGT	GCAGAGCCGG	GATATTTTCC	CTGTCCGTTT	CAACGGTCAC	CAATCCTGCA	420
CCGCTTTTTA	CAGCTGCTAA	AGCAGCCATG	ATGATGGCAC	CTCCATAAGG	ATAAGTACCA	480
CCAAGCAGCA	GCAGACGACC	ATAATCTCCT	TTATGACTTG	AACGAGAACG	TTCAATAATA	540
ACTTTTTCTA	GTAAGGTTTG	ATTAATCACT	TTCATCCTTT	TTCCCTCTCA	CTTTTATTAT	600
ACAACAAAAA	GGAGACGCAG	ACCTCCTTTT	GTAATCTTAT	ATCTAAAATT	TAATATTCAT	660
TTCTGCCATT	TTAGATATAG	CTATAGAAAA	TACACTCTAT	TAATCGAATG	TTTCTCTTAT	720
TTTCTATCCA	ATGTCCGAAG	TGCTGCTTGA	TAAGTTTGCT	CCATCAGCAT	GGTAATGGTC	780
ATAGGACCGA	CACCTCCAGG	GACTGGCGTG	ATATGGCTAG	CAAGTGGTGC	AACTGCCTCA	840
TAATCAACAT	CTCCACAGAG	CTTCCCATTT	TCATCTCGGT	TCATCCCAAC	GTCAATGACA	900
ACCGCACCTG	GTTTGACAAA	GTCAGCAGTC	ACAAACTTGG	CGCGGCCGAT	TGCGACTACA	960
AGAATATCTG	CTTTAGCAGC	CACCTTGGCA	AGATTATGAG	TTCGTGAGTG	GGCCAAGGTT	1020
ACTGTGCGAT	TTTTAGCCAA	AAGAAGCTGA	GCCATAGGTT	TTCCAACGAT	ATTTGAACGA	1080
CCGATTACGA	CCGCATTTTT	ACCTTCCAAG	TCAATCCCAT	ATTTCATGAA	CATTTCCATA	1140
ATTCCTGCAG	GTGTCGAGGG	AATCATGACT	GGATGTCCAG	ACCAAAGACG	TCCCATGTTT	1200
AGGGGATGGA	AACCATCCAC	ATCCTTTTCT	GGGTCAATGG	CTAATAAAAC	CGCCTCTTCA	1260
TCGATATGTT	TTGGTAATGG	CAACTGGACC	AAAATCCCAT	GCCAAGCTGG	ATCCTGATTA	1320
TATTTAGCAA	TCAGGTCTAA	CAATTCCTCT	TGAGTAATGG	TCTCTGGAAC	TCGCACTACT	1380
TCGGTACGGG	AACCAGCCGC	AAGAGCTGAC	CTCTCCTTGT	TGCGAACGTT	AAACTTGGCT	1440
GGCTGGATTA	TCCCCAACCA	AAATCACTAC	CAAACCAGGC	ACTAGAG		1487

(2) INFORMATION FOR SEQ ID NO:35:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1634 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

CGTGCCTTGG	CCAATGATCC	AAAAATCTTG	ATTTTCAGACG	AGTCGCTTCA	AATTTTCGGCC	60
CCTGGACCTT	TAAGACCAAC	CCAAGCAGAT	TTTGGCCCTT	GGTTGCAAGA	TTTGAACCAA	120
AAATTAGGCT	TGACTGTTGT	CCTGATTACG	CATGAAATGC	AGATTGTCAA	AGACATTGCC	180
AACCGTGTTG	CAGTTATGCA	GGATGGGCAT	TTGATTGAAG	AGAGTAGTGT	GCTTGAAATC	240
TTCTCAGACC	CTAAACAACC	TTTGA CTCAA	GACTTTATCT	CAACAGCTAC	AGGTATTGAC	300
GAAGCCATGG	TCAAAATCGA	GAAGCAAGAA	ATCGTGGAAC	ACTTGTCTGA	AAACAGTCTC	360
TTGGTGCAAC	TCAAGTACGC	TGGATCTTCA	ACAGACGAGC	CAC TTTTGAA	TGAATTGTAC	420
AAGCATTATC	AAGTAATGGC	TAATATTCTC	TATGGGAATA	TCGAAATCCT	CGATGGTACT	480
CCTGTTGGAG	AATTGGTGGT	GGTCTTGTC	GGTGAAAAAG	CAGCGCTGGC	AGGTGCTCAA	540
GAAGCCATTC	GTCAAGCAGG	CGTACAGTTA	AAAGTATTGA	AGGGAGGACA	GTAAGATGGA	600
ATCATTGATT	CAAACCTATT	TACCAAATGT	CTATAAGATG	GGTTGGTCTG	GTCAGGCAGG	660
CTGGGGAACA	GCTATCTACC	TAACCCTCTA	TATGACAGTT	CTTTCCTTCA	TTATCGGAGG	720
CTTCTTGGGG	CTAGTGGCAG	GTCTCTTTCT	CGTCTTGACA	GCGCCAGGTG	GTGTCTTGGA	780
GAATAAAGTC	GTATTCTGGA	TTTTAGACAA	AATTACCTCA	ATTTTTCGTG	CGGTTCCCTT	840
TATCATCCTC	TTGGCAATCT	TGTCACCACT	TTCTCACTTG	ATTGAAAAAA	CAAGTATCGG	900
GCCAAATGCA	AGCCCTTGTC	CCACTTTCTT	TTGCAGTCTT	TGCCTTCTTT	GCCCGTCAGG	960
TGCAGGTTGT	CTTGGCTGAA	ATGGATGGCG	GTGTCAATTGA	GGCGGGCTCA	AAGCGAGCGG	1020
AGCGACTTTC	TGGGACATCG	TGGGTGTTTA	CCTATCAGAA	GGTCTTCCAG	ATTTGATCCG	1080
TGTGACGACT	GTGACCTTGA	TTTCCCTTGT	TGGGGAAACA	GCTATGGCCG	GTGCGGTTGG	1140
AGCTGGTGGT	ATCGGTAACG	TAGCCATCGC	TTATGGATTT	AACCGCTACA	ATCACGATGT	1200
GACCATCTTG	GCAACCATCG	TTATCATTTT	GATTATCTTT	GCAATCCAAT	TCTTAGGAGA	1260
TTTCTTGACT	AAGAAAATTGA	GCCATAAATA	AAAAAGAGCC	GTGTGGCTCT	TTTTAACTGA	1320
TCAGATTTTC	TGGGCAAATT	TTTTACTCAA	GGCTTGTTCA	ATCAAGGCAC	CCACTAGGGC	1380
TCCGATGACA	ATACTTGCGA	TAAATAGAAG	GACAGTTCCA	GGGTTTGGAG	CGACCATGAT	1440
GCGGTCGATA	TATTCTTGGG	ATTTTCCTCT	TGCCAGAAGA	GTAGCCATAT	AGGCTTTGGG	1500
CGCAATCCAC	ATAAGCAAGA	TTGGTCCTGT	TGTACTAAAG	GCGAAAAATA	TGAAAGAAAG	1560
GAAGTTCTTT	GTTTTGTCCT	TGTATTTTCC	TAAATGAGCT	ACTCCATCTG	CTAGGAGGCC	1620
ACAGATAATT	CGAT					1634

(2) INFORMATION FOR SEQ ID NO:36:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1087 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

GGAATCATGA	TGATGTCACT	GCTAAATGGT	TTCTTAGAAA	AAATATTTCC	TGAGCGCTTA	60
CAGATTAGTT	TGGGCTTGCT	GATTTTATCA	TTGAGCGGTA	CAGCTCCCTT	CTGGTACCAA	120
GCCTATCCCT	TTGTCTTTGG	AACACGGCTT	CTCTTTGGTT	TGGGTCTTGG	GATGATCAAT	180
GCCAAGGCCA	TTTCTATTAT	CAGTGAACGC	TACCAAGGAA	AAAGGCGAAT	TCAGATGTTA	240
GGGCTACGCG	CTTCTGCAGA	GGTCGTTGGA	GCTTCTCTCA	TTACCTTGGC	CGTCGGTCAA	300
GTTGTTGGCC	TTTGGTTGGA	CAGCTATCTT	TCTAGCCTAT	AGTGCTGGAT	TTTTGGTGCT	360
GGCCCTTTAT	CTGCTCTTTG	TCCCTTATGG	AAAATCAAAG	AAAGAAGTCA	AGAAAAGAGC	420
GAAGGAAGCA	AGTCGTTTAA	CTCGAGAAAT	GAAAGGCTTG	ATTTTTACCT	TAGCTATCGA	480
AGCGGCAGTT	GTAGTTTGTA	CCAATACAGC	TATTACCATC	CGTATTCCAA	GTTTGATGGT	540
GGAAAGAGGA	TTGGGGGATG	CCCAGTTATC	TAGTTTTGTT	CTTAGTATCA	TGCAGTTGAT	600
CGGGATTGTG	GCTGGGGTGA	GTTTTTCTTT	CTTGATTTCT	ATCTTTAAAG	AGAAACTGCT	660
CCTCTGGTCT	GGTATTACCT	TTGGCTTGGG	GCAAATCGTG	ATTGCCCTTG	CTTCATCCTT	720
GTGGGTGGTA	GTAGCAGGAA	GTGTTCTGGC	TGGATTTGCC	TATAGTGTAG	TCTTGACGAC	780
GGTCTTTCAA	CTTGTCTCTG	AACGAATTCC	AGCTAAACTC	CTCAATCAAG	CAACTTCATT	840
TGCTGTATTA	GGCTGTAGTT	TCGGAGCCTT	TACGACCCCA	TTCGTTCTAG	GTGCAATTGG	900
CTTACTAACT	CACAATGGGA	TGTTGGTCTT	TAGTATCTTA	GGAGGTGGT	TGATTGTAAT	960
CTCTATCTTT	GTCATGTACC	TACTTCAGAA	GAGAGCTCTA	GGATTGATTC	CTAAGTTTTT	1020
CTTTTGATAC	TCAATGAAAA	TCAAAGAGCA	AACTATAGTT	GATTGAGTTT	GGAATAGTAT	1080
GCTGTAG						1087

(2) INFORMATION FOR SEQ ID NO:37:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1191 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

GGATTCCAAC	GATTATGAAC	TTGACTGGTC	CACTGATTCA	TCCAATGGCT	TTAGAAACAC	60
AGCTTTCTTG	GAATTAGTCG	TCCAGACTCC	TAGAAAGTAC	AGCTCAGGTT	TTGAAAATAT	120
GGTCGCAAAC	GTGCCATCGT	GGTTGCTGGA	CCAGAAGGGT	TGGATGAAGC	TGGCTTGAAC	180
GGAACAACCN	AGATTGCACT	TNNTGAAAAT	GGCGAAATCA	GCTTGTCAAG	CTTTACTCCA	240
GAGGATTTGG	GAATGGAAGG	CTATGCTATG	GAAGATATTC	GTGGTGGGAA	TGCTCAGGAA	300
AATGCAGAAA	TTTTGCTTAG	CGTTCTGAAA	AACGAAGCAA	GTCCATTCTT	GGAAACGACA	360
GTCTTGAATG	CTGGTCTTGG	TTTCTATGCT	AATGGTAAGA	TTGATAGCAT	CAAGGAAGGA	420
GTTGCCTTGG	CCCGTCAAGT	GATTGCTAGA	GGCAAGGCC	TTGAAAAACT	CAGACTGTTA	480
CAGGAGTACC	AAAAATGAGT	CAGGAATTTT	TAGCACGAAT	CTTAGAGCAG	AAGGCGCGTG	540
AGGTGGAGCA	GATGAAGCTG	GAGCAAATCC	AGCCTCTGCG	CCAGACCTAT	CGCTTGGCAG	600
AATTTTTGAA	GAATCATCAG	GACCGCTTGC	AGGTAATCGC	TGAGTCAAGA	AAGCTAGCCC	660

TAGTTTGGGA	GATATCAATC	TCGATGTGGA	TATTGTGCAA	CAGGCCCAGA	CTTATGAAGA	720
AAACGGAGCA	GTGATGATTT	CGGTGTTGAC	AGATGAGGTT	TTCTTTAAAG	GGCATTGGA	780
TTATCTACGG	GAAATTTCCA	GTCAGGTAGA	GATTCCGACG	CTCAACAAAG	ACTTTATCAT	840
AGATGAAAAG	CAAATCATCC	GCGCTCGCAA	TGCAGGTGCG	ACAGTTATCT	TGCTTATTGT	900
GGCAGCCTTG	TCCGAAGAAC	GCCTCAAGGA	ACTGTATGAC	TACGCGACAG	AGCTTGGTCT	960
GGAAGTCTTA	GTGGAGACTC	ACAATCTAGC	TGAACTAGAG	GTAGCCCACA	GACTTGGTGG	1020
CTGAGATTAT	CGGGGTCAAC	AACCGCAACT	TGACTACCTT	TGAAGTCGAC	TTGCAGACCA	1080
GTGTAGATTT	AGCCCCTTAC	TTTGAGGAAG	GTCGCTATTA	CATTTCTGAA	TCTGCCATTT	1140
TCACAGGGCA	GGATGCGGAA	CGACTAGCCC	CATACTTTAA	CGGAATTCCA	T	1191

(2) INFORMATION FOR SEQ ID NO:38:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 858 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

ATCGAATTTG	CCAACCAAGA	AAAATATCCC	TTGGATGGTT	CTTGCAATG	CAAGCAATAT	60
CATCGTTCGT	GATGGTGGGA	TTCGTGGATT	TGTCATCTTG	TGTGACAAGC	TCAATAACGT	120
TTCTGTTGAT	GGCTATACCA	TTGAAGCAGA	AGCTGGGGCT	AAC TTGATTG	AAACAAC TCG	180
CATTGCCCTC	CGTCATAGTT	TAAGTGGCTT	TGAGTTTGCT	TGTGGTATTC	CAGGAAGCGT	240
TGGCGGTGCT	GTCTTTATGA	ATGCGGGTGC	CTATGGTGGC	GAGATTGCTC	ACATCTTGCA	300
GTCTTGTAAG	GTCTTGACCA	AGGATGGAGA	AATCGAAACC	CTGTCTGCTA	AAGACTTGGC	360
TTTTTGTTAC	CGCCATT CAG	CTATTCAGGA	GTCTGGTGCA	GTTGTCTTGT	CAGTTAAATT	420
TGCCCTAGCT	CCAGGAACCC	ATCAGGTTAT	CAAGCAGGAA	ATGGACCGCT	TGACGCACCT	480
ACGTGAACTC	AAGCAACCTT	TGGAATACCC	ATCTTGTTGGC	TCGGTCTTTA	AGCGTCCAGT	540
CGGGCATT TT	GCAGGTCAGT	TCGAATTTCA	GAAGCTGGCT	TGAAAGGCTA	TCGTATCGGT	600
GGCGTAGAAG	TGTCAGAAAA	GCATGCAGGA	TTTATGATCA	ATGTCGCAGA	TGGAACGGCC	660
AAAGACTACG	AGGACTTGAT	CCAATCGGTT	ATCGAAAAAG	TCAAGGAACA	CTCAGGTATT	720
ACGCTTGAAA	GAGAAGTCCG	GATCTTGGGT	GAAAGCCTAT	CGGTAGCGAA	GATGTATGCA	780
GGTGGTTTTA	CTCCCTGCAA	GAGGTAGTGG	GGACCTGACA	GAGCCCCGAT	CGGTAAATCT	840
ATGAAAAAGA	AGGAATTT					858

(2) INFORMATION FOR SEQ ID NO:39:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 980 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

CTGAAAAAAC	AGGTTTTGAC	TATGNAGATT	GACAGACGAC	CGTTCGGAGG	TGCAGATATT	60
GATGCAGCAG	GACCTCCCTT	ACCTGATGAA	ACCCTTAAGG	CAAGTAGGGA	AGCAGATGCT	120
ATCCTACTAG	TAGCTATCGG	TAGTCCTCAG	TATGATGGAG	TAGCGGTTCC	CCCTGAACAA	180
GGCCTGATGG	CTCTCCGTAA	GAAC TCAATC	TTTACGCTAA	TATTCGTCCT	GTAAAAATCT	240
TTGACAGTCT	CAAGTATTTG	TCACCACTCA	AACCGGAACG	AATTTCTGGT	GTAGACTTCG	300
TCGTGGTGCG	TGAATTGACT	AGGCGAGATT	TACTTTGGAG	ATCATATCCT	TGAAGAGCGC	360
AAAGCGCGTG	ATATCAACGA	CTATAGCTAT	GAGGAAGTGG	AGCGGATTAT	TCGCAAAGCC	420
TTTGCCATCG	AATTGCAAGA	AATCGCAGAA	AAATCGTTAC	TAGTATCGAT	AAGCAAAATG	480
TTCTAGCGAC	CTCAAAACTC	TGGCGGAAAG	TAGCTGAGGA	AGTCGCACAG	GATTTCTCAG	540
ATGTAACCTT	GGAACACCAG	CTGGTAGACT	CAGCTGCTAT	GCTTATGATT	ACCAATCCTG	600
CTAAGTTTGA	TGTTATTGTA	ACGGAGAATC	TTTTTTGGAGA	TATTTTATCT	GATGAATCAA	660
GCGTCTTATC	TGGTACACTT	GGGGTTATGC	CATCAGCCAG	TCATTCTGAA	AATGGACCAA	720
GTCTCTATGA	ACCTATTCAC	GGTTCAGCAC	CTGATATTGC	AGGTCAAGGA	ATTGCCAATC	780
CTATTTCCAT	GATTTTATCA	GTTGTCATGA	TGTTGAGAGA	TAGTTTCGGA	CGTTATGAGG	840
ATACAGAGCG	TATCAAACGT	GCTGTTGAGA	CAAGTCTGGC	GGCAGGAATT	TTAACGAGAG	900
ATATAGGAGG	TCAGGCTTCA	ACAAAGGAAA	TGATGGAAGC	TATTATTGCA	AGGT TATGAA	960
GTTAGACGAA	AAAATTCGAT					980

(2) INFORMATION FOR SEQ ID NO:40:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 874 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

TCGATCTAGA	GAATTGCTCC	AGAGCTTCCT	GACCGTCCGC	TGCCTCAATA	GTTTCATAGC	60
CACAATCCGT	CAAATAATCA	CTGACCCCT	CACGGATCAT	CTCTTCATCT	TCTACAATTA	120
AAATTTTCAT	ACTTTAACTG	CTCTCTATTT	TTTATTTTTC	TTAGAATAAA	TACCTACTCT	180
ATTTTCTATT	ATAGTCTCTT	GCTGGCCTTT	TGTATGTAAG	CAACTGACCA	CTAGATAAAA	240
CGTTGTGAAA	TTCTTTTCTC	ATAAATTCCA	TAACTTTAGT	ATATTATATT	TAAGCACTAA	300
AGTACAAAGA	AAGCAACTGA	AAGCAATGAT	TTTCACCACT	GCTTTCAGAT	TTATTTTGA	360
TTGT TAAATA	GCTATT CCTA	TCCACTATTC	TTGAATAGAA	ACACAAGATG	CAATCTTTAT	420
TCCAGACTCA	TTTTTTAAAA	AATCAAATTT	ATTCACCATC	CAGCAAGAGC	TCTTTTGGTT	480

GTTCCTAAG	GAGATTGCTT	GAAGCAAGCG	CCATAACGAG	AACCACTAGA	ACCAAGGCAA	540
GGACAAAAAT	GATGATAAAG	TCTGATGTCT	GAATGGAAAT	GTCTAGGCTC	GACAAGGTCT	600
TGCTAAAGCC	ATCTACTTCT	GCACCGCCAC	CAAGGTTAGA	GGCTTGAGCC	GCCTTACTAG	660
CCTGTTTGGC	AACACCTGAA	GTCACATTGG	CAAGGACAGT	GTTTCCAATT	CGCACGGGCA	720
GTGTAATTAG	CTAGGAAGTA	AGCANAACT	AGAGCAGGGA	TAGCAATCAA	GATAGATTCG	780
GTGATGAATT	GACCCAAGAT	ACTTGCCCTG	TTGAGACCAA	TAGAGAGGAG	GATTCCCCACT	840
TCCTTGCCGA	CGGGCATTGA	TCCAAAGACT	GAGC			874

(2) INFORMATION FOR SEQ ID NO:41:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 762 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:

CTTGTAACGG	TCATAAAGTT	TCTGCAAACT	ACCATCCTTG	CTCCATTTAG	TAACCAAGTT	60
ATCAAGATAG	TCGTTGAGCT	CTGTATTTGA	TTTCTTGGA	ACAATACCGT	AGTCAGATGG	120
CTTGAAACTA	TCATCTAGTA	GTTCTGTGCG	TTTAACTAGT	GTAGCCAGAT	AGAATAGAGC	180
GGTCAACGGA	AAAGGCATCG	ATACGATGAG	CGTGAAGGGA	AGTAATCAAT	TCTGGGTAGG	240
AACCAAGTTC	GACGAATTTA	AACTTCAGAC	CTTCTTTTTT	ACCCAGTTCA	GTAATCAGGC	300
GTTGGGTGAT	AGAACCTTGG	GCGACTCCGA	TGGTTTTTGC	GTTTAGGTCC	TCAATCTTTT	360
TGATTTTGGC	AGATTTATTG	ACCAAAAATC	CAGAAGCGTC	TGTGTAGTAG	GGACTGGTAA	420
AGTTGTAGAG	TTTTTTTGC	TCGTCCGTGA	TGGTAAAGGT	CGCGATATCC	ATATCGACCT	480
GTTCAATTGTC	TAGAAGGGGG	CCGCGGGTTT	GTGCTGTAAC	CGGCACATAG	TGAATCTTGA	540
CCTTGAGTTC	ATCAGCTACC	ATTTTGGCCA	AGTCGGTTTC	GATACCAGAA	TAAGTACCGG	600
TCTTGGGATC	TTTGTTAACC	AAAATTGGGA	ACGTCTTGTT	TGACACCCGA	CAACCAGTTC	660
GCCTCTTTTT	TGAATGTCTG	CGATACTAGT	ATTAGCCTGG	ACTGGTTTGG	CAGCAACAAG	720
GCCGAAAAGG	CTAATCAATA	ATGCTGATAA	AAAGAATTTC	AT		762

(2) INFORMATION FOR SEQ ID NO:42:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1942 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:

CTCGAATTTT	TGGTGCTCCA	GAAACGGTTC	CAGCAGGAAG	CGTTGCTTTC	AAGGCATCCA	60
TGGCAGTGAG	TTCTGCAAGC	AAACGTCCCT	TGACCACACT	GGTCAAATGC	ATGACGTAGC	120
GGAAGAGCTC	CACCTCCATA	TACTTAGTAA	CTTGGACACT	GGCCGTTTCA	GAGATGCGGC	180
CAATATCGTT	ACGCCCCAAG	TCTACCAACA	TTCGATGTTC	TGCTGTTTCC	TTCTCATCAG	240
AGAGGAGGTC	AGTCGCCAAG	GCCTTGCTCT	CTCCATCCGT	AGCCCCCTCT	GGTCGCGTCC	300
CTGCAATCGG	ATTGGTTGTC	ACGATGCCAT	TTTTGACAGA	AACCAAACCT	TCTGGACTAG	360
CTCCGATGAT	TTGATAATCC	CCAAAATCAT	ACAAATAAAG	GTAATTAGAT	GGATTAGTCA	420
CGCGGAGATT	TCTGTAGAAG	TCAAATGGAT	TTCCAGTTAA	CTTCTGCGTG	AAGAAAACGC	480
TGGCTGAGTT	ACACATCGGA	ACATATCTCC	GTTACGAATC	AAGTCACGAG	CTGTTTCTAC	540
CATTCCCTCA	AACCTTATGTG	GAGCGATATG	CGGTTTGAAG	TCAAGTGGTG	ATAAATCCAA	600
GTCTTCAAAT	TCATTTGGAG	CAGGAATGCG	TAATTCCTCA	AGCACTTGGT	TCAAGGATTT	660
TTCCAAGGCC	TCTTGACTGC	GCTCACTATA	AAGTGCATCC	TCTATGACAT	GTTATCTTCT	720
CCTTCTTGTT	GGTCAAAGAC	CATATAGCTC	TCATAGACAA	AGAAATGCAT	GTCGGGCGTC	780
CCAATTGTAT	CCTCAGGGAT	TTGACCAATT	TCTTCATAAA	GCGAAATCAT	ATCGTAACCA	840
ACAAAACCAA	TGGCTCCCCC	ACCAAAAAGG	AGGTCTGAAT	GGTGCTGGCT	CTTATGAATC	900
ACTTCATAAA	GGAAATCCAA	GGGATCCCGA	TCAATCGCTT	GACCATTTTG	ATAGAGAACT	960
CCATTTTCAA	ACTTAATCTC	AAAAACTGGA	TTATAGGCTA	GGATAGAAAA	ACGAGCTGTT	1020
TCCTTGCTCT	TCGGAATACT	CTCTAAAATA	ACCTTATGTT	GCCCCTTTAA	GCGCATATAA	1080
GCCAAGATTG	GTGATAAGAC	ATCTCCATGA	ATGATTTCGT	CCATTGTCAT	TTCCCTTTCA	1140
GTTCTAATTC	GAGTTCGTGG	CGACTGTATG	AAAAATCCCC	ACGCAAAATA	ACTTGCGTGA	1200
GGACGAAATT	CGCGGTGCCA	CCTCAATTAT	AGGATTTCTC	CTATCTCTCA	TTCTGTCTC	1260
AGATATCTCC	TGTAACAGGC	TGTGCGATAA	AGGGCACTCC	CTTGAGAATG	ATGTTTTCTT	1320
CTCTCGTTTT	AGATGAACCC	AACTTTACAG	CTTCTCTGTC	TTGTTTTTCA	CAACCACAAG	1380
CTCTCTGTGA	GAGAAAAGAC	TGTAATTTTT	CCATCTATTA	TTTTTTTAGCT	TCTAGTAATC	1440
TGCAATCGCA	GCTAGGTCCT	TGCCTCCACG	ACCAGAGACA	TTGATGAAGA	GATGTTTCATC	1500
TCGGTACACC	TTTATACTCT	TCGAAAATCT	CTTCAAACCG	CGTCAACGTC	GCCTTGCCGT	1560
AGGTATGGTT	ACTGACTTCG	TCAGTTCTAT	CTGCAACCTC	AAAACAGTGT	TTTGAGCTGA	1620
CTTCGTCAGT	CTTATCGACA	ACCTCAAAAC	AGTGTTTTGA	GCAGCCTGCA	GCTAGTTTCC	1680
TAGTTTGCTC	TTTGATTTTC	ATTGAGTATT	ATTTTCAATTT	CTCCTGCAAT	TGAATTCTTG	1740
CTCAGCTTTT	TGTCTTCTAT	TTCTTTAAAA	TCAAAGTAGC	TCTTTTGTTA	ATAACTCGAT	1800
CAACAAACAT	CGTGGTACAA	GTATCTACTT	TGAAATTTAT	CAACCACTTA	ACAACGTATA	1860
CTGTATTTCT	AGGAAAACGA	TGACATTCTT	CCTAATAAAA	CTTCTCATAT	ATAGCATATA	1920
TTTCTACTCT	TTTTAATTCTG	AT				1942

(2) INFORMATION FOR SEQ ID NO:43:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1048 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:

CTGTTAAGAT	TGTTTCCGTG	CATCCACATA	GGATTTACCT	TGTCTGTATG	GGCCAATTCA	60
CCCATCAAAA	CGCCATAGGT	CTCATCTGTC	AAGATACTAG	ACATACCGAT	ATTGTACCAA	120
AGACTGGTAT	GACGGAAATA	AGTCGATGCG	TGTAAACTCA	ACAAAAAGAG	ACGCAAGTTG	180
ATTAGAAAAA	CCGTCATAGC	AATAGCTGCC	ACAGGAGCTT	GAACCACAAT	CAGTGCCAAC	240
ATGGCAAAC	GGGCACTCCC	AGCATAAACA	AAGAGACTCA	TCAAGCCCAT	CTCAACAGGT	300
GTCACATAGG	GCGCACCGAT	AGTCCCACAG	GCCAGGCCGA	TACTGACATA	GCCAAGAGCC	360
GTTGGCATGG	CTGCCTGCGC	CCCCTCCTAA	AATCCTTTTT	CTTTCATCTT	TCTCCTCATA	420
TTGTCTTAAT	AATACTCAAT	GAAAATCAAA	GAGCAAAC	GGAAATTAGC	CGCAGGNTGC	480
TCAAAACACC	GTTTTGAGGT	TGCAGATAGA	AACTGACGAA	GTCAGCTCAA	AACACCGTTT	540
TGAGGTTGCA	GATAGAACTG	ACGAAGTCAG	TAACATATAT	ACGGCAAGGC	GACGTTGACG	600
TGGTTTGAAG	AGATTTTCGA	AGAGTATTAG	AAAATGCCGA	TAAGGGTCTG	CATACCAAGG	660
CTGGTGAGGA	TGATGGCAAT	CCAGCAGACG	GCTCCGAGAA	CAATGGATTT	TCCACTGGAT	720
TTGACCATAG	CGACCAGATT	AGTTTTTGAGA	CCGATGGCAC	TCATGGCCAT	GATAATGAGG	780
AATTTAGAGA	GTTGTTTGAG	AGGGGTAAAG	AACTACTAG	ACACACCGAG	AGAGGTCAGA	840
AGGGTGGGTA	GGAGCGATGC	AAGGATGAAG	TAAAGGATAA	AAAGTGGGAA	GACTTTTTTTC	900
AGTTGTAAGC	CTTGCTTATT	TTTTTGCTCG	CGACTTTGCC	AGTAGGAGAG	AAAGAGAGTG	960
ATGGGGATGA	TAGCTAGGGT	GCGCGTGAGT	TTGACAATGG	TTGCGGATTC	GAGGGTATTG	1020
GTCTGGTAGA	GACTGTCCCA	AGCGCTAG				1048

(2) INFORMATION FOR SEQ ID NO:44:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1571 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:44:

AGAGCTGGTA	ATATTCCCAA	AGAAACGGCT	CAAATCGAAT	TAGAAAGCCT	TCTGCAAAAA	60
GGAATCCCAG	TCGCTCTGGT	ATCACGATGC	TTTAACGGTA	TTGCCGAGCC	TGTTTATGCC	120
TACCAGGGTG	GGGGCGTACA	GTTGCAAAAA	GCAGGCGTTT	TCTTTGTTAA	AGAACTCAAC	180
GCCCCAAAAG	CCCGCTTGAA	ACTCCTCATC	GCCCTCAATG	CCGGACTAAC	AGGACAGGCT	240
TTGAAAGACT	ATATGGAAGG	CTAATACTCT	TCGAAAATCT	CTGCAAACCA	CGTCAGCGTC	300
GCCTTACCGT	ATGTAGAGCA	CAAAATCAGG	AAATCTTCTC	GATTCCCTGA	TTTTTTCTAT	360
TTACGTTTTT	GTGTTGAGCT	ACGTTCTGTC	AAACCATGAG	GTAAGAGAAC	TTCACGTTCT	420
TCCAACCTTT	CCTTATGCAT	AATCTTGGTC	AACATACGCA	TACTAATGGC	ACCAAGGTCA	480
TAAAGAGGTT	GGGCAATCGT	TGTCAAGTTT	GGACGGGTAA	AGCGTGAGAT	TTGTGAATCA	540
TCACTAGTAA	TAATTCGATA	ATCTTCTGGC	ACAGAAACAC	CTTATCAGCC	AAACCGTTCA	600

AGACTCCTGC	TGCCAACTCA	TCACCTGTCA	CAACTGCTGC	AGTTGCATTT	GATGAAATCA	660
AACGCTCTGC	TAAGGCGTAA	CCATCATCAT	AGCTATATTT	AGATTCAAAT	ACCAAACCCT	720
CACTATAAGC	GATTCCCTGCT	TTTTTCAAGG	TTTCCTTGTA	GCCAACTAAA	CGAACCTTAC	780
CATTGATGTC	ATCCACTAGC	GGACCGCTAA	CGAAAGCAAT	ACGCTCATTT	TCTTTAGCAA	840
GGTAACTCAC	TGCATCAATT	GTTGCTTGCT	TATAGTCAAT	ATTGACACTT	GGCAACTGGT	900
GCTCAACATC	GACAGTTCCT	GCGAGAACAA	TCGGAGTACG	TGAACGCGAA	AATTCTGAGC	960
GAATTTTATC	TGTCAAGTGA	TAACCCATAT	AGATAATGCC	ATCTACCTGC	TTTGAAAAGA	1020
GGGTATTGAC	AACAGAAACT	TCTTCTCGT	TATCTTCATC	GCTATTAGCT	AGGACAATAT	1080
TGTACTTGTA	CATTTCTGCA	ATATCATCAA	TCCCCTTAGC	CAAACCTGAA	AAATAACCAT	1140
TGGTAATATT	TGGAATCACG	ACACCGACAG	TGGTTGTCTT	TTTACTTGCA	AGACCACGCG	1200
CAACTGCATT	TGGACGATAA	TCCAAACGAT	CAATTACCTC	TAGCACTTTT	TTACGGGTAT	1260
TCTCTTTTAC	ATTTTATTG	CCATTGACCA	CACGGCTGAC	CGTCGCCATG	GGAAACACCT	1320
GCTTCACGAG	CGACATCATA	AATGGTTACT	GTATCATCTG	CATTCATTCC	TTTTCTTGTC	1380
CTTTCTATCT	CCACACATTC	TTTACAAGT	AGAAGTGCTG	AATTGAAAGC	TCTATATCTT	1440
ACTTACAAAA	ATGAAGATGT	GAAAATTTTCG	TTTTCATATT	TCTACTTATT	CCATTCTATC	1500
ACTAATTGTA	AACACTTTCA	AGTGTTTTTT	GAAGATTGAT	TGAAAAAATT	TCATAGAAAA	1560
CCTAGGTTTA	G					1571

(2) INFORMATION FOR SEQ ID NO:45:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1682 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:45:

CTGACGTAAA	AAAGATTTTC	GGAAAAGTAT	CATCATCTAT	TTTAGACCAT	TTTCTTATAA	60
TAACCATTTT	ATTTTATTTT	GTCAAGGTCT	TTGAATTCTT	TCTTAAACAA	GCCTTGTAAT	120
CTCTACTTTT	GAAGAATTTA	TTTTTCCTTA	CTGACAAGAT	TTGAGACGGT	AGGAATCATT	180
GAAAATAACC	TAGCCAACAT	CAATCACAAT	CATTTCTCCT	TTCTCAATTA	CACTAAATTA	240
TAGTGTATTG	AATCTATAAC	AGTGCACCTT	GGCTGCTAAA	ATATTTCTAT	AAATTAATTT	300
GACTTTCCTG	ATAGAGTTGT	TCACATCTTA	TTTCAATTCA	CTATACTTTC	CCTTATACTC	360
AATGAAAATC	AAAGCGCAAA	CTAGGAAGCT	AGCCACAGGC	TGCTCAAAGC	ACTGCTTTGA	420
GGTTGTAGAT	AAGACTGACG	AAGTCAGTTA	CATATATCTA	CGGCAAGGCG	AAGCTGACGC	480
GGTTTGAAGA	GATTTTCGAA	GAGTATAAAG	TTTGTTTCTG	TATCTTTCAG	AAAAATAAGG	540
TATACTGTAT	GTAAACGATT	TCAAAGGAGT	CCAGTTATGG	CAAAAACATT	TTTTATTCCA	600
AATAAACAGA	GCATTTTAGG	AGAACAAGAG	ATTTTGAATG	CCAAGTCGAT	CTTGGCTATG	660
ATGTAGTCTA	TCTCCGTCAG	CCTCTTAATC	GTCTCGAGTA	TATTGAGTGT	GCGATAGTGG	720
GGCAATCACA	ATTTCTTTTT	AAGGTCAGTT	ATGCTGATGG	TCAAAAGGCT	TACCGTGTCTG	780
ATCTTCCTGA	CCTACTAACA	AAGACAGACT	GGCAGATTAT	CAAGTCATTT	TTAGATGTTT	840

TGCTTGCTTA	TACAGGGACT	GATATTGAAG	GGCTAGATGG	TTTTGATTTT	GAAGCTTATT	900
TCCAAGCAAG	TATTCAAGCC	TATCTAGCAG	ACCCTGTAGC	TCGTTTTACG	ATTTGCCAAC	960
GAATTTTTTAA	TCCTATTTTC	TTTAGTCGTG	AGAACTTGAA	AAGCTTTTTTA	GAGGCAGATG	1020
GCTTGCGCTCA	GTTTGAAGCG	CGTGTGCGTG	CGGTTCAAGA	GACAGATGCC	TACTTTGCGA	1080
GAGTTTCCTT	CTATCAGGAT	GGAGAAGGAA	AAGTGCATGG	CGTTTACCAT	CTAGCTCAAG	1140
GAGTCAAGAC	AGTTTTACCG	AGAGAACCGT	TTGTTCTCTG	AGCCTATATT	GAGCGAATTG	1200
GTGGATAAGG	AAGTCCAGTG	GGAGATTGAC	TTGGTTCAAA	TCACAGGAGA	CGGCTCTAAA	1260
CCAGAAGACT	ATGAATCCAT	AGCTCGCTTG	GACTATGCAA	AATTCTTAGA	GGTATTACCC	1320
CCATCTTTTTT	ACCACCAACT	AGACGCCAAT	CAAATAGAAA	TACAACCCAT	CCTAGGACAA	1380
GATTTTAAAA	CATTAGCACA	AGAAAAGTAA	AGCAGAAGCA	GGTCAATCGA	CTTGCTTTTTT	1440
TGACATAGAA	AAAATCCTGC	CAAGGATGAC	AGGATTGCTA	CTCAATGAAA	ATCAAAGAGC	1500
AAACTAGGAA	GCTAGCCGCA	GGCTGTACTT	GAGTACGGTA	AGGCGAAGCT	GACGTGGTTT	1560
GAATTTGATT	TTCGAAGAGT	ATGAATTTTA	AAGAAAGGCC	AAGATACGAA	GATAATCTCC	1620
AATCAGTGCC	ACTTCAGCTT	CCAAGAAGAA	GAAGATTATA	ACTCCCGTTC	CCCAAGGACA	1680
GA						1682

(2) INFORMATION FOR SEQ ID NO:46:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3041 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:46:

ATCGAATTAA	AAATGAGGTA	TTCAGGCTTG	TGATTTTCTA	TGGAAGTTAA	TAGTGATTGC	60
CTCTAATGCT	TACAAGTGAT	ATTAAAAATA	GAGGACCTAG	TGATGTCAAT	CATTTCAACT	120
GATTTAACCC	CTTTTCAAAT	AGATGATACA	TTGAAAGCAG	CCTTGCGAGA	AGATGTTTCA	180
TCCGAAGATT	ACAGTACCAA	TGCCATTTTT	GATCATCATG	GCCAAGCCAA	GGTGTGCGTT	240
TTTGCCAAGG	AAGCTGGTGT	TTTAGCGGGG	CTAACCGTTT	TTCAAAGGGT	TTTTACCCTA	300
TTTGATGCCG	AGGTGACCTT	CCAGAATCCT	CATCAATTTA	AGGATGGGGA	TCGTTTGACT	360
AGTGGCGATT	TGGTTTTAGA	AATCATAGGC	TCGGTGAGAA	GTCTCTTAAC	ATGTGAACGC	420
GTTGCCTTGA	ATTTTTTACA	ACATTTATCA	GGGATCGCTT	CGATGACAGC	TGCTTATGTA	480
GAAGCCTTAG	GCGATGATTG	CATTAAGGTA	TTTGATACTC	GAAAAACTAC	TCCTAATTTA	540
CGTCTTTTTG	AGAAATATGC	CGTGAGAGTT	GGCGGTGGCT	ATAATCATCG	CTTTAATTTA	600
TCAGATGCTA	TCCTGCTAAA	AGACAATCAC	ATTGCGGCAG	TAGGTAGTGT	TCAAAGGGCA	660
ATTGCTCAAG	CGCGTGCCTA	TGCTCCTTTT	GTGAAAATGG	TCGAGGTGGA	AGTGGAAGC	720
CTTGCTGCTG	CCGAAGAAGC	TGCGGCGGCG	GGTGCTGATA	TTATCATGTT	GGATAATATG	780
TCATTGGAAC	AGATTGAACA	GGCCATTACC	CTAATTGCAG	GACGTTCTCG	GATTGAATGT	840
TCTGGAAATA	TTGATATGAC	CACTATTAGC	CGTTTTTCGTG	GTTTAGCGAT	TGATTACGTC	900
TCCAGTGGTA	GTTTAACCCA	TAGTGCTAAG	AGTCTTGATT	TTTCCATGAA	GGGTTTAACC	960

TACCTTGATG	TCTAAGTTGT	AAAATAAACT	AACTTTTTTAA	AGGATGTCTT	TCCTCTAGAA	1020
CGAGTTTTAT	GTCAGATAGT	TTAAACGCCT	CTTCAAATAT	AGTAAAATGA	ACCAAAAATA	1080
GTACACAATG	TGGTATAATC	TTCTTATGGC	ATATTCAATA	GATTTTCGTA	AAAAAGTTCT	1140
TTCTTATTGT	GAGCGAACAG	GTAGTATAAC	AGAAGCATCA	CACGTTTTCC	AAATCTCACG	1200
TAATACCATT	TATGGCTGGT	TAAAGCTAAA	AGAGAAAACA	GGAGAGCTAA	ACCACCAAGT	1260
AAAAGGAACA	AAACCAAGAA	AAGTTGATAG	AGATAGACTT	AAAAACTATC	TTACTGACAA	1320
TCCAGACGCT	TATTTGACTG	AAATAGCTTC	TGAATTTGGC	TGTCATCCAA	CTACCATCCA	1380
CTATGCGCTC	AAAGCTATGG	GCTACACTCG	AAAAAAGGAC	CACACCTACT	ATGAACAAGA	1440
CCCAGAAAAA	GTAGCCTTAT	TTCTTAAAAA	TTTTAATAGT	TTAAAGCACC	TAGCACCTGT	1500
TTAGATTGAT	GAAACAGGAT	TCGATACTTA	TTTTTATCGA	GAATATGGTC	GCTCATTAATA	1560
AGGTCAGTTA	ATAAGAGGTA	AAGTATCTGG	AAGAAGATAT	CAGAGGATTT	CTTTGGTTGC	1620
AGGTCTAACA	AATGGTGAGT	TAATCGCTCC	AATGACTTAC	GAAGAGACGA	TGACGAGCGA	1680
CTTTTTTGAA	GCATGGTTTC	AGAAGTTTCT	CTTACCAACA	TTAACCACAC	CATCGGTTAT	1740
TATTATGGAT	AATGCAAGAT	TCCATAGAAT	GGGTAAGTTA	GAACTTTTAT	GCGAGGAGTT	1800
TGGGCATAAA	CTTTTACCTC	TTCCCTCCCTA	CTCGCCTGAG	TACAATCTTA	TTGAGAAAAC	1860
ATGGGCTCAT	ATCAAAAAGC	ACCTCAAAAA	GGTATTACCA	AGTTGCAATA	CCTTTTATGA	1920
GGCTCTTTTG	TCCTGCTCTT	GTTTCAATTG	ACTATAGTTC	ACGGATACAG	TTGGGAAAGA	1980
AGTTAAATGT	AGTTGGATTT	CCACTAAAGG	TTGATGAGTA	AGTTTTTGTA	TCTGAACCTG	2040
ATTGGCCGCA	AGCAGCTAAA	AGCAAAGCAG	ATGCAAAAGT	CAGACCTGCA	CCAAGGACAC	2100
GCTTCTTTAT	GTTTCATCTC	TTTCTCCTTA	ATAGTGGGAA	TTTGTAAGT	TAATTGAATT	2160
TCAAGAATGA	AGGTTTTATA	AACTTTGGTT	ATAAAAAACA	AAGGATTTCT	GTCTTTTATA	2220
CAGTCCTCCC	CTTGTTTTTA	TACGATTTCA	ATTTTAAATT	TTTCTGCAAA	AAATATTTAT	2280
AGTAATTCCA	CACAGAAAGC	ATCCCATGGA	ACTAAGATTT	GTTTTTCAAA	GACTTCTTGA	2340
GCTAGGGTGT	TTTCAATCAA	GACAGATTTG	ACTTTTCCTT	CTACTGTCAA	GTCTTGCTCT	2400
TCATTGGACA	AGTTAGCCAC	AACTAGGAAG	CGACGGTCGC	CATCCTTACG	TATATAAGCA	2460
AAGACCTTAT	CAGCCGTATC	AAGCAATTCA	AAGTCAGCTC	GAATTAGCCA	ACTATTCTCC	2520
TTGCGAATTT	GGACCAGTTT	CTGATAGGTA	TAGAAAATAG	AATCTGGATT	TGCCAGCGCT	2580
TC TTGGACGT	TGATCATCTC	GTAATTTGGA	TTAACTGCCA	ACCAAGGTTG	ACCTGTTGAG	2640
AAACCAGCGT	TTTTGCTCTC	GTCCCATTCG	ATAGGGGTAC	GGGCATTGTC	ACGTCCAATA	2700
ACACGGATAC	TGTCCATGAT	TTCTTGTCATC	GGAACACCTT	TTTCAAGAGC	CTCACGCGCA	2760
TAGTTGAGAG	ATTCAATATC	TTCTACTTGA	TCCAGTGTTT	CAAACGGATA	GTTGGTCATC	2820
CCAATCTCCT	CACCTTGGTA	GATATAAGGA	GTTCCTCTCA	TAAGATGAAG	CAAGATTGCA	2880
AAGGCTTTGG	CAGATTTTTT	GCGGTATTCT	TGGTCATTTT	CCCAGATTGA	GACAATACGA	2940
GGGAGGTCAT	GGTTGTTCCA	GAAGAGGGAA	TTCCAGCCGT	CCTCAACTCC	TAACTCTGTC	3000
TGCCATTTGT	TGAAGATTTT	TTTTAACTTA	GCGATATTCA	G		3041

(2) INFORMATION FOR SEQ ID NO:47:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 4694 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:47:

TTAATTTAAA	TTCTTAAAAT	TTTTTCATAA	TAATCTCCCT	ATAAAAATAA	AGTCGCCCAA	60
TCAGGCGGCT	TATTTTTTTG	AAAAATGGGC	TTGGTGCCTG	AGAATAAATA	GCTTAGTGAT	120
AGAAGAAAAT	GGGGAAATAT	GGTATAATGA	AACGATAGAT	TTTTGAATAG	GAATAAGATC	180
ATGTTTGGAT	TTTTTAAGAA	AGATAAAGGC	TGTGGAAGTA	GAGGTTC CGA	CACAGGTTCC	240
TGCTCATATC	GGCATCATCA	TGGATGGCAA	TGGCCGTTGG	GCTAAAAAAC	GTATGCAACC	300
GCGAGTTTTT	GGACATAAGG	CGGGCATGGA	AGCATTGCAA	ACCGTGACCA	AGGCAGCCAA	360
CAAACCTGGG	GTCAAGGTTA	TTACGGTCTA	TGCTTTTTTCT	ACGGAAAACT	GGACCCGTCC	420
AGATCAGGAA	GTCAAGTTTA	TCATGAACCT	GCCAGTAGAG	TTTTATGATA	ATTATGTCCC	480
GGAACCTACAT	GCGAATAATG	TTAAGATTCA	AATGATTGGG	GAGACAGACC	GCCTGCC TAA	540
GCAAACCTTC	GAAGCTTTAA	CCAAGGCTGA	GGAATTGACT	AAGAACAACA	CAGGATTGAT	600
TCTTAATTTT	GCTCTTAACT	ATGGTGGACG	TGCTGAGATT	ACACAGGCGC	TTAAGTTGAT	660
TTCCCAGGAT	GTTTTAGATG	CCAAAATCAA	CCCAGGTGAC	ATCACAGAGG	AATTGATTGG	720
TAACATCTCT	TTTACCCAGC	ATTTGCCTAA	GGACTTACGA	GACCCAGACT	TGATTATCCG	780
TACTAGTGGA	GAATTGCGTT	TGAGCAATTT	CCTTCCATGG	CAGGGAGCCT	ATAGTGAGCT	840
TTATTTTACG	GACACCTTAT	GGCCTGATTT	TGACGAAGCG	GCCTTG CAGG	AAGCTATTCT	900
TGCCTATAAT	CGTCGCCATC	GCCGATTTGG	AGGAGTTTAG	GAGGAAATAT	GACCCAGGAT	960
TTACAGAAAA	GAACCTTGTT	ATGCAGGGAT	TGCCCTGACT	ATTTTCCTAC	CAATTTTAAT	1020
GATTGGGGGC	TCTTGCTTCA	GATAGCAATC	GGAATCATAN	CCATGCTAGC	CATGCATGAA	1080
CTTTTGAAGA	TGAGAGGTCT	AGAGACCATG	ACGATGGAGG	CCTCTTGACC	CTCTTTGCAC	1140
NTTNGTATTG	ACCATTCCCC	TGGAATCGAA	TTACCTGACT	TTTTTGGCCAG	TTGATGGGAA	1200
TGTGGTTGCC	TATAGTGTTT	TGATTTCAAT	CATGTTAGGA	ACGACCGTTT	TTAGCAAGTC	1260
TTATACGATT	GAGGATGCGG	PTTCCCCTCT	TGCTATGAGC	TTCTACGTGG	GCTTTGGATT	1320
TAATGCTTTA	CTAGATGCTC	GTGTTGCAGG	TTTGGACAAG	GCTCTCTTAG	CCTTGTGTAT	1380
CGTCTGGGCG	ACAGACAGTG	GTGCCATATCT	TGTTGGGATG	AACTATGGGA	AACGAAAGTT	1440
AGCACCAAGG	GTATCGCCTA	ATAAAACCCT	TGAGGGTGCC	TTGGGTGGTA	TTTTAGGAGC	1500
AATTTTAGTA	ACCATTATCT	TTATGATAGT	TGACAGTACA	GTTGCTCTTC	CATATGGAAT	1560
TTACAAGATG	TCAGTCTTTG	CTATTTTCTT	TAGCATTGCT	GGACAATTTG	GTGATTTACT	1620
AGAAAGTTCT	ATCAAACGTC	ATTTTGGTGT	TAAGGATTCT	GGGAAATTTA	TCCCTGGACA	1680
TGGTGGTGTT	TTGGATCGTT	TCGATAGTAT	GTTGCTTGTA	TTTCCAATCA	TGCAC TTATT	1740
TGGACTCTTT	TAATCAAAAAG	ACGGAGGAAA	CGCTATGCTC	GGAATTTTAA	CCTTTATTCT	1800
GGTTTTTGGG	ATTATTGTAG	TGGTGCACGA	GTTGCGGCAC	TTCTACTTTG	CCAAGAAATC	1860
AGGGATTTTA	GTACGTGAAT	TTGCCATCGG	TATGGGACCT	AAAATCTTTG	CTCACATTGG	1920
CAAGGATGGA	ACGGCCTATA	CCATTGCAAT	CTTGCCCTCTG	GGTGGCTATG	TCCGCATGGC	1980
CGGTTGGGGT	GATGATACAA	CTGAAATCAA	GACAGGAACG	CCTGTTAGTT	TGACACTTGC	2040
TGATGATGGT	AAGGT TAAAC	GCATCAATCT	CTCAGGTAAA	AAATTGGATC	AAACAGCCCT	2100
CCCTATGCAG	GTGACCCAGT	TTGATTTTGA	AGACAAGCTC	TTTATCAAAG	GATTGGTTCT	2160
GGAAGAAGAA	AAAACATTTG	CAGTGGATCA	CGATGCAACG	GTTGTGGAAG	CAGATGGTAC	2220
TGAGGTTCTG	ATTGCACCTT	TAGATGTTCA	ATATCAAAAT	GCGACTTTAT	CTGGGGCAAA	2280
CTGATTACCA	ATTTTGCAGG	TCCTATGAAC	AATTTTATCT	TAGGTGTTGT	TGTTTTTTGG	2340

GTTTTAATCT	TTATGCAGGG	TGGTGTGAGA	GATGTTGATA	CCAATCAGTT	CCATATCATG	2400
CCCCAAGGTG	CCTTGGCCAA	GGTAGGAGTA	CCAGAAACGG	CACAAATTAC	CAAGATTGGC	2460
TCACATGAGG	TTAGCAACTG	GGAAAGCTTG	ATCCAAGCTG	TGGAAACAGA	AACCAAAGAT	2520
AAGACGGCAC	CGACTTTGGA	TGTGACTATT	TCTGAAAAGG	GGAGTGACAA	ACAAGTCACT	2580
GTTACACCCG	AAGATAGTCA	AGGTCGTTAC	CTTCTAGGTG	TTCAACCGGG	GGTTAAGTCA	2640
GATTTTCTAT	CCATGTTTGT	AGGTGGTTTT	ACAACTGCTG	CTGACTCAGC	TCTCCGAATT	2700
CTCTCAGCTC	TGAAAAATCT	GATTTTCCAA	CCGGATTGTA	ACAAGTTGGG	TGGACCTGTT	2760
GCTATCTTTA	AGGCAAGTAG	TGATGCTGCT	AAAAATGGAA	TTGAGAATAT	TCTTGTAATT	2820
CTTGGCAATG	ATTTCCATCA	ATATTGGGAT	TTTAAATCTT	ATTCCGATTC	CAGCCTTGGA	2880
TGGTGGTAAG	ATTGTGCTCA	ATATCCTAGA	AGCCATCCGC	CGCAAACCAT	TGAAACAAGA	2940
AATTGAAACC	TATGTCACCT	TGGCCGGAGT	GGTCATCATG	GTTGTCTTGA	TGATTGCTGT	3000
GACTTGGAAT	GACATTATGC	GACTCTTTTT	TAGATAATCG	AGGAATATTA	TGAAACAAAG	3060
TAAAAATGCCT	ATCCCAACGC	TTCGCGAAAT	GCCAAGCGAT	GCTCAAGTTA	TCAGCCATGC	3120
TCTTATGTTG	CGTGCTGGTT	ATGTTTCGCCA	AGTTTCAGCA	GGTGTTTATT	CTTATCTACC	3180
ACTTGCCAAC	CGTGTGATTG	AAAAAGCTAA	AAACATCATG	CGCCAAGAAT	TCGAAAAGAT	3240
TGGTGCTGTT	GAGATGTTGG	CTCCAGCCCT	TCTTAGTGCA	GAATTGTGGC	GTGAATCAGG	3300
TCGTTACGAA	ACCTATGGTG	AAGACCTTTA	CAAACTGAAA	AACCGTGAAA	AATCAGACTT	3360
TATCTTAGGT	CCAACTCACG	AAGAAACCTT	TACAGCTATT	GTCCGTGATT	CTGTTAAATC	3420
TTACAAGCAA	TTGCCACTCA	ACCTTTATCA	AATTCAGCCC	AAGTATCGTG	ATGAAAAACG	3480
CCCACGTAAT	GGACTTCTTC	GTACACGTGA	GTTTATCATG	AAGGATGCTT	ATAGTTTCCA	3540
CGCTAACTAT	GATAGTTTGG	ATAGTGTTTA	TGATGAGTAC	AAAGCAGCCT	ATGAGCGTAT	3600
TTTCACTCGT	AGTGGTTTAG	ACTTCAAGGC	TATTATTGGT	GACGGTGGAG	CCATGGGTGG	3660
TAAGGATAGC	CAAGAATTTA	TGGCCATTAC	ATCTGCTCGT	ACAGACCTTG	ACCGCTGGGT	3720
TGTCTTGGAC	AAGTCAGTTG	CCTCATTTGA	CGAAATTCCT	GCAGAAGTGC	AAGAAGAAAT	3780
CAAGGCAGAA	TTGCTCAAAT	GGATAGTCTC	TGGTGAAGAT	ACCATTGCTT	ACTCAAGTGA	3840
GTCTAGCTAT	GCAGCTAACT	TAGAAATGGC	AACAAACGAG	TACAAACCAA	GCAACCGTGT	3900
TGTCGCTGAA	GAAGAAGTTA	CTCGTGTTGA	AACGCCAGAT	GTTAAATCAA	TTGATGAAGT	3960
TGCAGCCTTC	CTCAATGTTC	CAGAAGAACA	AACGATTAAA	ACCCTCTTCT	ACATTGCAGA	4020
TGGTGAGCTT	GTTGCAGCCC	TTCTAGTTGG	AAATGACCAA	CTCAACGAAG	TCAAGTTGAA	4080
AAATCACTTG	GGAGCAAATT	TCTTTGACGT	TGCTAGCGAA	GAAGAAGTGG	CGAATGTTGT	4140
TCAAGCAGGA	TTTGGTTTAC	TTGGACCAGT	TGGTTTGCCA	GAGAATATTA	AAATTATTGC	4200
AGATCGTAAG	GTGCAAGATG	TTCGCAATGC	AGTTGTCGGT	GCTAACGAAG	ATGGCTACCA	4260
CTTGACTGGT	GTGAACCCAG	GCCGTGATTT	TACTGCAGAA	TATGTGGATA	TCCGTGAAGT	4320
TCGTGAGGGT	GAAATTTCCC	CAGATGGACA	AGGTGTCCTT	AACTTTGCGC	GTGGTATTGA	4380
GATCGGTCAT	ATTTTCAAAC	TCGGAACCTG	CTATTTCAGCA	AGCATGGGAG	CAGATGTCTT	4440
GGATGAAAAAT	GGTCGTGCTG	TGCCAATCAT	CATGGGATGT	TACGGTATCG	GTGTCAGCCG	4500
TCTTCTTTCA	GCAGTGATGG	AGCAACACGC	TCGCCCTCTT	GTTAACAAAA	CGCCAAAAGG	4560
TGAATACCGT	TACGCTTGGG	GAATCAATTT	CCCTAAAGAA	TTGGCACCAT	TTGATGTGCA	4620
TTTGATTACT	GTTAATGTCA	AGGATGAAGA	AGCGCAAGCC	TTGACAGAAA	AACTTGAAGC	4680
AAGCTTGATG	GGAG					4694

(2) INFORMATION FOR SEQ ID NO:48:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1352 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:48:

CTCGTAAGTT	CGGAAGCTAT	CTACACAAGA	AATTAACCGC	TGCCTAAAGG	AGAAGCCATG	60
TCAACATATA	ACTGGGATGA	GAAGCATATC	CTTACCTTTC	CTGAAGAAAA	AGTAGCCCTT	120
TCTACTAAGG	ATGTCCATGT	TTACTATGGT	AAAAATGAAT	CCATTAAGGG	GATTGATATG	180
CAATTTGAAA	GAAATAAAAT	TACAGCTTTG	ATTGGTCCGT	CGGGATCGGG	GAAATCTACC	240
TACTTACGCA	GTCTCAATCG	CATGAATGAT	ACCATTGATA	TTGCTAAAGT	AAC TGGGCAG	300
ATTCTCTATC	GTGGAATTGA	TGTCAACCGT	CCAGAAATCA	ACGTTTATGA	AATGCGTAAA	360
CACATTGGAA	TGGTTTTTCA	ACGCCCAAT	CCATTTGCTA	AATCGAATTT	ACCGTAATAT	420
TACCTTTGCG	CATGAACGTG	CTGGAGTTAA	GGATAAGCAA	GTCCTAGATG	AAATCGTAGA	480
AACCTCCCTT	AGTCAGGCTG	CCCTTTGGGA	TCAGGTAAA	GACGATCTCC	ACAAGTCAGC	540
CTTGACCTTA	TCAGGTGGTC	AGCAACAACG	TCTCTGTATC	GCTCGTGCCA	TCTCTGTAA	600
GCCAGATATC	CTCTTAATGG	ATGAGCCAGC	CTCAGCCTTG	GATCCGATTG	CGACCATGCA	660
ACTAGAAGAG	ACCATGTTTG	AGCTCAAGAA	AACTTTTACC	ATCATCATTG	TAACGCATAA	720
TATGCAGCAG	GCTGCTCGTG	CAAGTGACTA	TACAGGCTTC	TTTTACTTGG	GTGATTTGAT	780
TGAGTATGAC	AAGACTGCAA	CTATTTTCCA	AAATGCCAAG	CTACAGTCCA	CCAATGACTA	840
TGTATCTGGT	CACTTTGGTT	AGAAAGGAAA	CCGTATGACA	GATGCGATTT	TACAGGTATC	900
AGACCTGTCC	GTTTATTATA	ATAAAAAGAA	GGCTTTGAAT	AGTGTTTCCC	TATCTTTCCA	960
ACCTAAGGAA	ATTACAGCCT	TGATTGGTCC	ATCTGGATCA	GGGAAGTCAA	CCCTCCTCAA	1020
GTCTCTCAAC	CGCATGGGAG	ATCTCAATCC	AGAGGTGACC	ACAACTGGAT	CCGTGGTGTA	1080
CAATGGTCAC	AACATCTACA	GTCCGCGTAC	AGATACGGTT	GAATTACGTA	AGGAAATCGG	1140
AATGGTTTTT	CAACAACCTA	ATCCTTTCCC	TATGACTATC	TATGAGAATG	TTGTCTACGG	1200
GCTTCGTATC	AATGGAATTA	AGGATAAGCA	GGTTCTGGAT	GAAGCCGTAG	AAAAAGCCTT	1260
GCAAGGTGCC	TCTATCTGGG	ATGAGGTCAA	GGATCGTCTA	TATGATTTCAG	CTATTGGATT	1320
GTCAGGTGGT	CAACAGCAGC	GTGTCTGCGT	GG			1352

(2) INFORMATION FOR SEQ ID NO:49:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2258 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:49:

AACTTCGACC	GTGATAAACA	AGCTGAGCTT	TGACATACTT	GTAGCCAACC	TAAAAGCCGT	60
TCTTCAAGGC	CTCAAACCAG	CTGCAACTCA	TTCAGGAAGC	CTGGATGAAA	ATGAAGTGGC	120
TGCCAATGTT	GAAACCAGAC	CAGAACTCAT	CACAAGAACT	GAAGAAATTC	CATTTGAAGT	180
TATCAAGAAA	GAAAATCCTA	ATCCCAGCTG	GTCAGGAAAT	ATTATCACAG	CAGGAGTCAA	240
AGGTGAACGA	ACTCATTACA	TCTCTGTACT	CACTGAAAAAT	GGAAAAACAA	CAGAAACAGT	300
CCTTGATAGC	CAGGTAACCA	AAGAAGTTAT	AAACCAAGTG	GTTGAAGTTG	GCGCTCCTGT	360
AACTCACAAAG	GGTGATGAAA	GTGGTCTTGC	ACCAACTACT	GAGGTAAAAAC	CTAGACTGGA	420
TATCCAAGAA	GAAGAAATTC	CATTTACCAC	AGTGACTCGT	GAAAATCCAC	TCTTACTCAA	480
AGGAAAAACA	CAAGTCATTA	CTAAGGGTGT	CAATGGACAT	CGTAGCAACT	TCTACTCTGT	540
GAGCACTTCT	GCCGATGGTA	AGGAAGTGAA	AACACTTGTA	AATAGTGTCG	TAGCACAGGA	600
AGCCGTTACT	CAAATAGTCG	AAGTCGGAAC	TATGGTAACA	CATGTAGGCG	ATGAAAACGG	660
ACAAGCCGCT	ATTGCTGAAG	AAAAACCAA	ACTAGAAATC	CTAAGCCAAC	CAGCTCCTGC	720
TGAGGAAAGC	AAAGCTCTTC	CTCAAGATCC	AGCTCCTGTG	GTAATAGAGA	AAAAACTTCC	780
TGAAACAGGA	ACTCACGATT	CTGCAGGGAC	TAGTAGTCGC	AGGACTCATG	GCCACACTAG	840
CAGCCTATGG	ACTCACTAAA	AGAAAAGAAG	ACTAAGTCTT	TTCGATAAAA	AATAAACAGC	900
GAGATTGAAG	CTCGCTGTTT	ATTTTTTAAT	TAATCACCTA	GTCCAAGACG	TTCAAAGATA	960
TCATCCACTC	GTTTGGTGTA	ATAAACTGGG	TTGAAGATTT	CATCGATTTC	TTCTTGTGTG	1020
AGACGTGATG	TTACTTCTGA	ATCTGCCTCA	AGAAGTGGTT	TAAAGTCTAC	TTGGTTGTCC	1080
CAAGAGTAGG	CTGTTTTTGG	TTGCACCAAG	TCATAGGCTT	GCTCACGGGT	CATGCCTTTT	1140
TCAATCAATG	TCAACATAGC	CCGTTGGCTA	AAGATAAGAC	CAAAAGTCGA	GTTTCATGTTT	1200
CGGATCATAT	TTTCTGGGAA	GACTGTCAAG	TTCTTGACGA	TATTTCCAAA	ACGGTTGAGC	1260
ATGTAGTCAA	TCAAAATGGT	CGTATCTGGT	GTGATGATAC	GCTCAGCTGA	TGAGTGAGAA	1320
ATATCGCGTT	CGTGCCAGAG	AGCGACGTTT	TCATAAGCCG	TAATCATGTG	ACCACGAATG	1380
ACACGCGCCA	GACCAGTCAT	ATTTTCAGAA	CCGATTGGGT	TGCGTTTGTG	AGGCATTGCT	1440
GAAGACCCTT	TTTGCCCTTT	AGCAAAGAAC	TCTTCTACTT	CGCGTTGCTC	AGATTTTTTGT	1500
AGACCACGAA	TCTCAGTCGC	CATACGTTCT	ATTGAAGTCG	CAATGCTGGC	AAGAACCGCA	1560
AAGTACTCAG	CGTGAAGGTC	ACGAGGAAGG	ACTTGTGTTA	AAGATTCCCTT	GGGCACGGAT	1620
GCCAAGATTT	ATCGCAGACA	TACTCCTCTA	CAAATGGTGG	GATATTGGCA	AAGTTCCCAA	1680
CCGCACCAGA	AATCTTACCA	GCTTCTACAC	CAGCAGCCGC	ATGCTCGAAG	CGCTCGATAT	1740
TGCGTTTCAT	TTGCTGTAC	CAAGTTGCTA	ATTTAAGACC	AAAGGTTGTC	GGCTCAGCGT	1800
GCACACCATG	AGTACGCCCC	ATCATGATGG	TGAACTTGTG	CTCCTTGGCC	TTGTCAGCGA	1860
TGATATTAGT	GAAGTTTTC	AGGTCACGAC	GGATGATGTC	GTTGGCCTGC	TTGTAGAGGT	1920
AACCATAAGC	AGTATCCACC	ACGTCGGTAG	AAGTTAACCC	ATAGTGAACC	CACTTGCGCT	1980
CTTCACCAAG	AGTCTCAGAA	ACCGCACGCG	TGAAAGCCAC	CACATCGTGG	CGCGTCTCCT	2040
GCTCAATTTT	CAAAATACGG	TCGATGTCAA	AGTCCGCCTT	CTTGCGAATC	AAAGCCACAT	2100
CTTCCTTAGG	GATTTCCCCC	AACTCAGCCC	ATGCCTCGTC	AGAGAGGATT	TCCACCTCAA	2160
GCCAAGCACG	GTATTTATTT	TCTTCACTCC	AAATATTCGC	CATCTCAGGG	CGAGAGTAAC	2220
GGTTGATCAT	GTGTTAATTT	TTCTTTCTT	CTTAAGAT			2258

(2) INFORMATION FOR SEQ ID NO:50:

(i) SEQUENCE CHARACTERISTICS:

162

- (A) LENGTH: 4392 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:50:

CCCTTTTGCC	TCTCCCTTTG	GTGCAGATTC	TTTTGGGAAT	TGTGATTGGT	CTCTTTTAC	60
CCAATACTGA	CTTTCATCTT	AATACGGAGT	TGTTTTTGCC	CTGGTTATCG	GACCCTTGCT	120
TTTCCGAGAG	GCTGAAGAAG	CAGATGTTAC	GGCTATTTTA	AAACACTGGC	GAATCATTGT	180
TTATCTCATA	TTTCCAGTGA	TTTTTATCTC	GACCCTGAGT	TTGGGTGGCT	TGGCCCATCT	240
TCTTTGGTTC	AGCCTTCCCT	TGGCAGCTTG	CTTGGCTGTT	GGGGCAGCCC	TTGGTCCTAC	300
GGACTTGGTG	GCCTTTGCCT	CTCTTTCGGA	GCGTTTTAGC	TTTCCTAAGC	GCGTGTCCAA	360
TATTCTTAAG	GGCGAAGGAC	TCTTGAATGA	TGCTTCTGGT	TTGGTGGCTT	TTCAGGTAGC	420
TTTGACAGCT	TGGACAACTG	GAGCTTTTTT	TCTGGGGCAA	GCTAGCAGTT	CGCTCATCTT	480
TTCAATCCTA	GGCGGTTTTT	TAATTGGATT	TTTAACAGCC	ATGACCAACC	GCTTCCTCCA	540
TACCTTCTTG	CTAAGTGTGC	GCGCAACGGA	TATTGCCAGT	GAACTTTTAT	TAGAATTCTGA	600
GTTTGCCTCT	AGTGACCTTC	TTTCTGGCAG	AAGAAGTCCA	TGTTTCAGGG	ATTATTGCCG	660
TCGTAGTTGA	TCGAATTTTA	AAGGCAAGTC	GCTTCAAGAA	AATCACGCTC	CTCGAAGCCC	720
AAGTGATAC	GGTGACCGAG	ACGGTCTGGC	ATACAGTGAC	CTTTATGCTC	AACGGTTCCTG	780
TCTTTGTGAT	TTTAGGGATG	GAGTTGGAAA	TGATAGCAGA	ACCTATCTTG	ACCAATCCAA	840
TCTATAATCC	TCTACTTTTA	TTGCTATCTC	TCATCGCCCT	TACCTTTGTC	CTCTTTGTCA	900
TTCGTTTTAT	TATGATCTAT	GGCTATTATG	CCTATAGAAC	CCGACGCCCTA	AAGAAAAAGC	960
TAAATAAGTA	TATGAAGGAC	ATGTTTCTCT	TGACCTTTTC	AGGTGTTAAG	GGAACGGTGT	1020
CGATTGCTAC	GATTCTCTTG	ATACCAAGTA	ATCTAGAACA	GGAGTATCCT	CTCTTGCTTT	1080
TCCTTGTTGC	AGGTGTGACG	CTTGTCAGCT	TTTTAACAGG	TCTCTTGGTC	TTGCCTCATC	1140
TTTCTGATGA	AGAGGAAGAA	AGCAAGGATT	ATCTCATGCA	TATCGCCATT	TTGAATGAAG	1200
TAACGCTAGA	GTTGGAAAAA	GAGTTGGAAG	ACACCAGAAA	TAAACTTCCC	CTCTATGCGG	1260
CTATTGACAA	TTCGATCATG	GACGTATTGA	AAATCTCATT	TTAAGCCAAG	AAAACCAGGA	1320
TGATCAAGAA	GACTGGGCTG	CTTTGAAAAT	CGAATTCTTA	GTATTGAAAAG	TGATGGTTTG	1380
GAACAGGCCT	ATGAAGAGGG	GAACATTAGC	AATCGTGCTT	ACCGAGTTTA	CCAACGTTAT	1440
CTGAAAAATA	TAGAACAAGG	AATCAATCGT	AACTTGCCCT	CAAGACTGAC	CTATTATTTT	1500
CTTGTTTCCT	TGAGGATTTT	ACGTTTTCTT	CTTCATGAAG	TTTTTACTCT	TGGAAAGACC	1560
TTCCGTAGCT	GGAAGGACAA	GGAGCAAAGC	CGTCTCCGTG	CTCTTGATTA	TGACCAAATT	1620
GCAGAGCTCT	ATCTTGCCAA	TACAGAGATG	ATTATTGAAA	GTTTGGAAAA	CCTGAAGGGA	1680
GTCTACAGAC	GCTCTTTGAT	TAGTTTTATG	CAGGAGTCTC	GTCTTCGAGA	AACAGCTATT	1740
ATCAGCAGTG	GTGCCTTTGT	CGAACGGGTT	ATCAATCGTG	TCAAACCCAA	CAATATCGAT	1800
GAAATGCTGA	GAGGCTATTA	TCTGGAGCGC	AAGTTGATTT	TGAATACGA	AGAAAAACGA	1860
TTGATTACGA	CTAAGTATGC	CAAGAAATTA	CGACAAAATG	TAAATAACTT	AGAGAACTAT	1920
TCCTTGAAGG	AAGCTGCCAA	TACCCTGCCG	TATGATATGG	TGGAATTGGT	AAGAAGAAAT	1980
TAGTTAATAC	TCTTCGAAAA	TCTCTTCAAA	CCACGTCAGC	GTGCGCTTGG	ATTATATATG	2040

TGACTGACTT	CGTCAGTTTC	ATCTACAACC	TCAAAGCAGG	GCTTTGAGCA	ACCTGCGGCT	2100
AGCTTCCTAG	TTTGCTCTTT	GATTTTCATT	GAGTATAAGA	TTGTAAGTGA	AGGAGTGTGA	2160
CATGAAAAAA	TGGGGAAAAGA	GCCTGAACTA	GTCCTGTCTA	CTTTTACCCA	ATCACACTTC	2220
CATTTGGTAC	AGCTGGATCA	ACTGTGAGAA	GGGATCGAAT	TTGCCATCAT	G TTCAGCTGA	2280
GAGAATCATA	CCCTGGCTGA	CATATTTTTT	CATCATTTTA	CGTGGTTTGA	GGTTAGCAAC	2340
GATTTGAACT	TTCTTGCCGA	CCAATTCTTG	TTCATTTTGA	TAGTATTTTG	CAATTCCTGA	2400
AAGAATCTGA	CGATCTTCTC	CATCACCAGC	ATCCAAGCGG	AATTGAAGCA	ACTTATCTGA	2460
ACCTTCTACT	TTAGACACTT	CTTTGACTTC	TGCGACACGG	ATTTCAACCT	TGTCAAAGTC	2520
TTCAAACCTG	ATTTTCATCCT	TGTTTAGTTT	GAGCTCAACT	TCGTCCGGAT	TCCATTCCTT	2580
TTCGACTGCT	GGTTTATTGC	CTTCCATTTG	TTCTTTGATA	TAGGCGATTT	CTTCTTCCAT	2640
ATTTAGACGT	GGAAAGATAG	GTGTTCCTTT	GGCAACTACA	GTCACATCTG	CTGGGAAGTC	2700
AGCCAAACTC	AAGTTTTTCAA	GACTAGAAAC	TTCTTCCAAA	CCAAGTTGAG	TCAAAACCTGC	2760
ACGACTAGTT	TCCATCATAA	ATGGTTCAAT	CAAGTGAGCA	ACTACACGAA	TGCTGGCTGC	2820
CAAGTGGCTC	ATGACACTTG	CCAATTGGTC	ACGAAGAGCT	TCATCCTTGT	CCAAGACCCA	2880
TGGTGCAGTC	TCATCGATGT	ATTTATTGGT	ACGAGAGATC	AGAGTCCAGA	CTGCTTCAAG	2940
CGCACGTGGA	TAGTCAACTG	CTTCCATGTG	TGTATGGAAG	TCTGCGATTG	ATTTTTCTGTC	3000
AACCTCAGCA	AGAACATGAT	CAAATTCAGT	CACACCTTCT	ACATAGGCAG	GGATTTGTCC	3060
ATCAAAGTAC	TTATTAATCA	TGGAAACCGT	ACGGTTAAGG	AGGTTCCTAA	GGTCATTAGC	3120
CAATTCATAG	TTGATACGAC	CGACATAGTC	TTCAGGAGTA	AAGGTTCCGT	CTGAACCAAC	3180
TGGAAGGTTA	CGCATGAGGT	AGTAACGAAG	TGGATCTAGT	CCATAACGCT	CTACCAACAT	3240
TTCAGGGTAA	ACGACATTCC	CTTTTGACTT	AGACATTTTT	CCGTCTTTCA	TGACAAACCA	3300
ACCATGGGCA	ATCAAACGAT	CAGGTAATTT	AACATCCAAC	ATCATAAGAA	GGATTGGCCA	3360
G TAGATAGAG	TGGAAGCGAA	GGATGTCTTT	TCCTACCATA	TGGAAGACTG	TTCCATTCCA	3420
GAAC TTGTCA	AAGTTACCAT	GTTCTGTCTG	AGCGTAGCCA	AAAGCTGTCTG	CATAGTTAAG	3480
AAGGGCATCA	ATCCAAACGT	AGACAACGTG	TTTTGGATTT	GATGGGACAG	GCACTCCCCA	3540
TGTAAAGGTT	GTACGAGATA	CCGCCAAATC	TTCCAAACCT	GGCTCGATGA	AGTTGCGTAG	3600
CATTTTCATTA	AGACGACCAT	CTGGCGTGAT	AAATTCAGGA	TGAGCTTTGA	AAAATTCGAC	3660
CAAACGGTCT	TGGTATTTGC	TAAGGCGAAG	GAAGTATGAT	TCTTCAGAAA	CCCATTCAAC	3720
CTCATGACCT	GATGGAGCAA	TACCACCAGT	CACATTTCCA	GCTTCATCAC	GGAAAACTTC	3780
TGCCAGCTGG	CTTTCTGTAA	AGAATTCTTC	GTCTGATACT	GAATACCAAC	CAGAGTATTC	3840
ACCCAAGTAG	ATATCATCTT	GAGCAAGTAA	GCGTTCAAAG	ACCTGTGCGA	CAACTTTTTTC	3900
ATGGTAGTCA	TCGGTTGTAC	GGATAAATTT	ATCGTATGAG	ATATCTAGTA	ATTGCCAGAG	3960
TTCTTTAACT	CCAACCGCCA	TTCCATCAAC	ATAGGCTTGA	GGTGTAATAC	CAGATTCGAA	4020
TTCCGCTTTC	TGCTGGATTT	TCTGACCATG	TTCATCAAGA	CCTGTCAGAT	AAAATACATC	4080
G TAGCCCATC	AGGCGTTTGT	AACGTGCTAG	GACATCACAT	GCGATAGTTG	TGTAGGCAGA	4140
ACCGATATGA	AGTTTCCCAG	ATGGATAGTA	AATCGGCGTT	GTAATATAAA	AATTTTTTTTC	4200
AGACATAATT	TTTCCTTTCC	AGGCAAATGA	AACCTGTTTT	TCTAACACTT	CATTATATCA	4260
CATTTTTTAAT	GAATTTTCGAT	AGGGAAATCC	ATACCAAAAC	AAGATAGACG	AGTGTCCATC	4320
TTGTTGATCT	CATTCATAAC	GAAGGGCTTC	AATTGGATCA	AGTTTCGATG	CCTTGTTGGC	4380
TGGCAAGACT	CC					4392

(2) INFORMATION FOR SEQ ID NO:51:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1941 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:51:

AATTAGTATT	CTCAACCTTT	TTATCTTGAT	AGTTCAAGAT	GGCATTCGTT	GAATTGGTAA	60
CATAGTAACT	ATCCACTCCC	TTCAGTTTAG	CTGCCTCTTG	AACCCAGGAT	TCTTGCGGTT	120
TTGGCGGTTT	AACAGGAATT	CTTTTCTTTT	TCCAGAAACC	GTAAAAGCTG	ATTGTTTCTG	180
AGTAAAAGAC	CCATCTTTAC	TTTTTTTAGG	AGAGAAAAAG	ACGCTAATAT	TTTCTGAGA	240
TTTAGTCATA	TCTTTATTGA	CTTGACGAGA	TAGGGAATCA	CCCAAAGCCA	TAATCACAAC	300
AACTGATGAA	ACACCGATAA	TAATCCCAAT	CATAGTAAGC	AAAGAACGCA	TCTTGTGAGC	360
CATGATAGAT	GAAAAGGCAA	ATTTCAGATT	CTGCATCTTA	GTTTTCTCTC	TTTCCTAACT	420
GAGCACTGTC	AGACGAAATG	ACCCCATCCC	GAATGACAAT	CTGACGTTTG	GCATAGGCAG	480
CAATCTCAGG	CTTCATGCGT	TACCATGATA	ATGGTTTTTC	CTTCTTTATT	CAAATCAACC	540
AATAATTGCA	TAATTTGGTT	ACCTGTTTTG	GTATCCAAGG	CTCCTGTCGG	TTCATCCGCT	600
AGGATAATAG	AAGGATTGTT	TACCAAGGCA	CGCGCAATGG	CTACACGTTG	CTTTTGACCA	660
CCAGATAATT	CTGAAGGTAA	ATGGTGACTA	CGTTCATCA	ATTCAACCTT	GTCTAAATAT	720
TCCTCAGCCA	ACTTGCGACG	TTTTGAAGAC	GAAACTCCTG	CGTAAATCAA	GGGCAATTCT	780
ACATTTTGCA	GAGCATTGAG	CTTCGATAGA	AGAAAGAACT	GCTGAAAGAC	AAAACCGATT	840
TGTTGGTTAC	GGACCTTAGC	TAGTTGTTTT	TCACCAAGCC	CAGCCACTTC	TTGACCTTCA	900
AGATAATATT	CTCCACTGGT	TGGTGTATCC	AACATGCCAA	TCGTATTCAT	CAGAGTGGAC	960
TTACCAGACC	CAGATGGTCC	CATGATGGCT	ACAAATTCAC	CCTCATTCAC	TTCTAGATTG	1020
ATATTTTGA	GAACCTGCAG	TTCTTGGTCA	CCATTACGGT	AACCTCTGAA	GATATTTTTT	1080
AGACTAATTA	GTTGCTTCAT	CAGCCTTCAC	CTCTTTTCCT	TCTTCCAAGG	AAGATGTTGG	1140
ATTACTGATG	ACCTTAGCAC	CGTTCGTAA	ACCAGAAGTG	ATTTCTTGAT	TTTCTGCGTC	1200
AGCATTTCCC	AATGAAACCT	CAACTTTTTT	AGCCTTTTGT	TGTTTCATCCA	CAATCCAGAC	1260
ATAATTTTGA	CTATCATCCA	TTACTAGACT	GCTAACAGGA	ACAAGAATAG	CCTTAGTTTT	1320
GCTTTTAACC	TCAATGTTGA	CAGAAAAACC	TTGTTTCAA	TCACCAACCT	CGCCTGTCAC	1380
ATCAATAGTA	TAAGGGTATT	TAGAACCTGT	ATTATTCCCG	GCTGCTGGAC	TAGCTGCTTC	1440
ACCATTGTTT	TTAGGATAGT	CAGAAATATA	GGCTTAATTT	CCCAGTCCAT	TTTTTATCAG	1500
GATACACTTT	AGAAGTAAAG	CTTACTTCTT	GACCTACAGA	AAGGTTGGCT	AGATTGTACT	1560
CAGACAATTC	TCCCTTGACT	TGTAAATTTT	CATTGCTGAC	AATATGAACC	ATAACTTGAC	1620
TCGCCCCGTG	TGGAGATTTA	GAAACATTGC	TATTGACTTC	GACTACAGTT	CCCTCTAGGG	1680
TACTGAGAAC	AGTTGTTGCA	TCCAATTGAC	TTTGAGCCTT	GCTTAATTGC	GCTGCAGCAT	1740
CTGCACGCGC	ATCACGGGCA	TCACCCAATT	GAGCATCAAT	AGAAGCAACA	GAATTTCCAG	1800
CCACTGGAGT	TGGGCTTTGC	ACCGTTGCAT	CTTCTCCTCC	TACTGGCGCT	GGTAACTGTG	1860
GAGCCTGAGC	TGAAGCGGCT	TCATTTCTGT	CTTGATTGAG	TTCATTGATA	TGACGATCTG	1920
CCTTAGCTAC	TGCTCGACTA	G				1941

(2) INFORMATION FOR SEQ ID NO:52:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1335 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:52:

ATCGAATTCC	CTATTTTAAAC	ACTTTCCTTTT	CTAAAACAGT	CTATATTTTAA	TTTCAAACCTG	60
TATTATATTT	TTGAAAAAAT	AAAGTCCTTT	TTTCTTTTTT	TCAGAAAAAA	GGGTATAATA	120
AAAGAAAATA	AGCAGTAACA	CTCAATGGAA	ATCGAAAAAG	CAAACCTAGGA	AGCTAGCCGC	180
AGATTGCTCA	AAACACTGTT	TTGAGGTTGC	AGATAGAGCT	GACGTGGTTT	GAAGAGATTT	240
TCGAAGAGTA	TAAAAAGGTG	CTAGGCATGT	TGATTTTTCC	TTTGTTAAAT	GATTTGTCAA	300
GAAAAATCAT	CCATATTGGA	CATGGATGCC	TTTTTTGCTG	CAGTGGAAT	CAGGGATAAT	360
CCTAAACTCA	GAGGAAAACC	TGTCATTATT	GGAAGCGACC	CTCGGCAAAC	AGGTGGACGG	420
GGAGTCGTTT	CTACCTGTAG	TTATGAGGCA	AGAGCTTTTG	GTGTCCATTC	TGCCATGAGT	480
TCCAAGGAAG	CTTATGAACG	TTGTCCCCAG	GCTGTCTTTA	TCTCAGGGAA	TTCGATGAGA	540
AATACAAGTC	TGTGGGACTC	CAGATTCGAG	CTATTTTTAA	GCGCTATACA	GATTTGATTG	600
AACCCATGAG	CATTGACGAA	GCCTATTTGG	ATGTGACAGA	AAATAAACTC	GGTATCAAGT	660
CAGCGGTCAA	AATTGCTCGC	CTCATTCAAA	AAGATATCTG	GCAAGAACTC	CATCTAACTG	720
CTTCCGCAGG	CGTTTCTTAC	AACAAATTCT	TAGCTAAAAT	GGCGAGTGAT	TATCAAAAAC	780
CACATGGTTT	GACAGTGATT	CTACCTGAAC	AGGCTGAGGA	TTTTCTCAAA	CAAATGGATA	840
TTTCCAAATT	TCATGGAGTA	GGAAAAAGA	CAGTAGAACG	TCTTCATCAA	ATGGGCGTTT	900
TTACTGGTGC	TGATTTACTT	GAAGTTCCTG	AGGTAACCTT	AATAGACCGT	TTTGGTAGAC	960
TAGGCTATGA	TCTGTATCGA	AAGGCTCGTG	GCATTCACAA	CTCTCCAGTC	AAATCCAATC	1020
ACATCCGTAA	ATCAATCGGC	AAGGAGAAAA	CCTACGGGAA	GATTCTCCGT	GCTGAGGAAG	1080
ATATCAAAAA	AGAGAGCTGA	CTCTTCTATC	AGAAAAAGTC	GCTCTCAATC	TACATCAACA	1140
AGAAAAAGCT	GGAAAAATTG	TCATTTTGAA	AATCCGCTAC	GAGGACTTTT	CAACTCTTAC	1200
CAAACGAAAA	AGTATTGCTC	AAAAAACACA	AGATGCTAGT	CAGATAAGCC	AAATAGCCCT	1260
GCAACTCTAT	GAAGAATTAA	GTGAGAAAGA	AAGAGGTGTC	CGCCTATTGG	GGATTACCAT	1320
GACTGGATTT	TAAAG					1335

(2) INFORMATION FOR SEQ ID NO:53:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1796 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:53:

TCCAAGCTAG	CTATTTTCGTG	GAAGGGGCTT	CGGTTGGCAG	AACCTGGTGA	ATTTACCCAA	60
ACGTGCTTTT	TTAAACGGTC	GCGTAGACTT	GACACAGGCA	GAGGCTGTGA	TGGATATCAT	120
CCGTGCCAAG	ACTGACAAGG	CCATGAACAT	TGCGGTCAAA	CAATTAGACG	GCTCCCTTTC	180
TGACCTCATT	AACAATACCC	GTCAAGAAAT	CCTCAATACA	CTTGCCCAAG	TTGAGGTCAA	240
TATCGACTAT	CCTGAATATG	ATGATGTTGA	GGAAGCTACT	ACTGCCGTTG	TCCGTGAGAA	300
GACTATGGAG	TTTGAGCAAT	TGCTAACCAA	GCTCCTTAGG	ACAGCACGTC	GTGGTAAAAT	360
CCTTCGTGAA	GGAATTTCAA	CGGCTATCAT	TGGACGTCCC	AACGTTGGGA	AATCAAGCCT	420
TCTCAACAAC	CTCTTGCGTG	AGGACAAGGC	TATCGTAACC	GATATCGCTG	GGACAACACG	480
AGATGTCATC	GAAGAGTACG	TCAACATCAA	TGGTGTTCCCT	CTAAAAATTGA	TTGACACAGC	540
TGGTATTCGT	GAAACGGATG	ATATCGTTGA	ACAAATCGGT	GTTGAGCGTT	CGAAAAAGC	600
CCTCAAGGAA	GCCGACTTGG	TTCTACTAGT	GCTAAATGCC	AGTGAACCAC	TGACTGCGCA	660
AGACAGACAA	CTTCTTGAAA	TTAGCCAAGA	TACCAATCGC	ATTATTCTAC	TTAATAAAAC	720
CGACCTGCCA	GAAACGATTG	AAACTTCGAA	ACTACCTGAA	GACGTTATCC	GTATTTCACT	780
CCTTAAAAAC	CAAAACATCG	ACAAGATTGA	AGAGCGAATC	AACAACCTCT	TCTTTGAAAA	840
TGCTGGCTTG	GTGAGCAAG	ATGCTACTTA	CTTGTCAAAC	GCCCCTCACA	TTTCCCTGAT	900
TGAAAAAGCA	GTTGAAAGCC	TACAAGCCGT	TAATCAAGGT	CTTGAGCTGG	GGATGCCAGT	960
TGATTTGCTT	CAAGTTGACT	TGACTCGTAC	TTGGGAAATC	CTCGGAGAAA	TCACTGGGGA	1020
TGCTGCTCCA	GATGAACTCA	TCACCCAAC	CTTTAGCCAA	TTCTGTTTAG	GAAAAATAAGA	1080
AAAATCCATG	ATCCTTCATT	CGGTCATGGA	TTTTATTGTC	TTTATTAGTA	ATCTGGTCTT	1140
AAGACCCCTG	TTACAGTTGC	CTTAGTTGCT	TCGTAGTCGC	CATCTACGAC	AACCTTGATA	1200
ATGCGTTTGA	CATCTTCTTC	TGGTGCTGGA	ACAAGAGGTA	GACGAGTGGG	TCCAGCTTCA	1260
AATCCCATAT	AGTTAAGAAT	TGCCTTAAC	GGAGCAGGAC	TTGGATAAGA	GAAGAGAGCA	1320
TTAACCTTAG	GAATGAATTT	ACGCTGAATT	GCTGCGGCTT	TCTTCATATC	GCTTTCTGCA	1380
ATGGCAGTAA	ACATCTCGTG	CATTTTCATC	CCATTTGTAT	GAGAGGCAAC	AGAAATAACC	1440
CCATCCGCCC	CAAGGTTTCAT	GGCATGGAAA	GCATCTCCAT	CCTCACCTGT	ATAAATCAAG	1500
AACTCTTCAG	GCTTGCTGCTC	AATCAAGTAA	GCCATATTAG	CCAAGCTAGT	ACATTCTTTG	1560
ACACCGATAA	TATTTGGATG	GTCAGCCAAG	CGAAGCATGG	TTTCTGGAGT	CAATTTCGACA	1620
ACTACACGCC	CTGGAATGTT	ATAGATAATA	ATTGGTAGGT	CAGAAGCATC	TGCAATAGCC	1680
TTAAAGTGCT	GATACATCCC	TTCTTGAGAA	GGTTTGTTGT	AGTAAGGAAC	AATAGCAAGC	1740
CCAGCTGCGA	AACCACCAA	TTCCGCTACT	TCTTTGACAA	ACTCAATAGA	GTCACG	1796

(2) INFORMATION FOR SEQ ID NO:54:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2337 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:54:

CTTCGTACAG	GTGGTTCCTA	TGCAAGGGTG	GAAGCCAATC	GTCAGAACAA	CAAGCATCTT	60
CATCAAGCCA	GAAGTGGAGC	AATTACAAAA	AGAAATTGCT	GAAGAAGAAG	CAAGCTTGGG	120
TTCAGAAGAA	GTGGCTTTGA	AGACCTTGCA	AGATGAGATG	GCCAGATTGA	CCGAGTCATT	180
AGAAGCTATT	AAATCTCAAG	GAGAGCAGGC	ACGTATTTCAG	GAGCAAGGCT	TGTCCCTCGC	240
TTATCAGCAA	ACTAGTCAGC	AAGTTGAAGA	ACTGGAAACT	CTTTGGAAAC	TCCAAGAAGA	300
GGAAATAGAT	CGTCTTTCCG	AGGGAGATTG	GCAAGCGGAT	AAGGAAAAAT	GCCAAGAGCG	360
TCTTGCTGCA	ATCGCCAGTG	ACAAGCAAAA	TCTGGAAGCT	GAGATTGAAG	AGATTAAGTC	420
TAATAAAAAAT	GCCATCCAAG	AACGCTATCA	AAACTTGCAG	GAAGAGCTAG	CGCAAGCTCG	480
TTTGCTTAAG	ACAGAACTGC	AAGGGCAAAA	ACGTTATGAA	ATTGCTGATA	TTGAACGCTT	540
AGGCAAGGAA	TTGGACAATC	TTGATTTTGA	ACAAGAGGAA	ATCCAGCGCC	TTCTTCAAGA	600
AAAGGTTGAC	AATCTTGAGA	AGGTTGATAC	AGAATTGCTC	AGTCAACAGG	CGGAAGAATC	660
CAAACTCAG	AAAACGAACC	TCCAACAAGG	TTTGATTTCGC	AAACAGTTTG	AGTTGGATGA	720
TATAGAAGGT	CAGCTGGATG	ATATTGCTAG	TCATTTGGAT	CAGGCTCGCC	AGCAGAATGA	780
GGAGTGGATT	CGCAAGCAAA	CACGTGCTGA	AGCTAAGAAA	GAAAAGGTCA	GCGAGCGCTT	840
TGCCGCCATC	TACAAAGTCA	ATTAACAGAC	CAGTACCAGA	TTAGCCATAC	TGAAGCTCTA	900
GAAAAAGCGC	ATGAATTGGA	AAACCTCAAT	CTGGCAGAGC	AAGAAAGTTAA	GGATTTAGAG	960
AAGGCTATTC	GCTCACTGGG	TCCTGTCAAT	ATAGAAGCTA	TTGACCGGTA	CGAAGAAGTT	1020
CACAACCGTC	TGGACTTTCT	AAATAGTCAG	CGAGATGATA	TTTTGTTCAGC	GAAAAATCTG	1080
CTCCTTGAAA	CCATTACAAA	GATGAATGAT	GAGGTTAAGG	AACGCTTTAA	ATCAACCTTT	1140
GAAGCTATTC	GTGAGTCCTT	TAAAGTGACC	TTCAAGCAGA	TGTTTGGCGG	AGGTCAGGCA	1200
GAAGCTATTC	GTGAGTCCTT	TAAAGTGACC	TTCAAGCAGA	TGTTTGGCGG	AGGTCAGGCA	1260
CAGGTAAGAA	AATCCAGTCG	CTTAACCTCA	TGAGTGGTGG	TGAAAAAGCC	CTATCGGCTC	1320
TTGCCTTGCT	TTTCTCCATT	ATTCGTGTCA	AGACCATTCC	TTTTGTTCATC	TTGGATGAGG	1380
TGGAAGCTGC	GTTGGATGAA	GCCAATGTTA	AACGTTTTTG	GGATTACCTC	AACCGCTTTG	1440
ACAAGGACAG	CCAGTTTATC	GTCGTAACCC	ACCGTAAGGG	AACCATGGCA	GCGGCCGATT	1500
CCATCTATGG	AGTGACCATG	CAAGAATCGG	GTGTTTCAAA	GATTGTTTCA	GTTAAGTTAA	1560
AAGATTTAGA	AAGTATTGAA	GGATGACAAT	TAAACTAGTA	GCAACGGATA	TGGACGGAAC	1620
CTTCCTAGAT	GAGAATGGGC	GCTTTGATAT	GGACCGCCTC	AAGTCTCTCT	TGGTTTCCTA	1680
CAAGGAAAAA	GGGATTTACT	TTGCGGTGGC	TTGCGGTGGC	GGATTTCTGT	CTCTGGAAAT	1740
CGAATTATTT	GCTGGTGTTT	GTGATGACAT	TATTTTTCATC	GCGGAAAATG	GCAGTTTGGT	1800
AGAGTATCAA	GGTCAGGACT	TGTATGAAGC	GACTATGTCT	CGTGACTTTT	ATCTGGCAAC	1860
TTTTGAAAAG	CTGAAAACGT	CACCTTATAT	AGATATCAAT	AAACTGCTCT	TGACGGGTAA	1920
GAAGGGTTCA	TATGTTCTAG	ATACGGTTGA	TGAGACCTAT	TTGAAAGTGA	GTCAGCATTA	1980
TAATGAAAAT	ATCCAAAAAG	TAGCGAGTTT	GGAAGATATC	ACAGATGACA	TTTTCAAATT	2040
TACAACCAAC	TTCACAGAAG	AAACGCTAGA	AGCTGGTGAA	GCTTGGGTCA	ATGATAATGT	2100
CCCTGGTGTC	AAGGCTATGA	CAACTGGCTT	TGAATCTATT	GATATTGTTC	TGGACTATGT	2160
CGATAAGGGT	GTAGCTATTG	TTGAATTAGC	TAAAAAACTT	GGCATCACAA	TGGATCAGGT	2220
CATGGCTTTT	GGAGACAATC	TTAATGACTT	ACATATGATG	CAGGTTGTGG	GACATCCTGT	2280
AGCTCCTGAA	AATGCACGAC	CAGAGATTTT	AGAATTAGCA	TAAGACTGTG	ATTGGTCT	2337

(2) INFORMATION FOR SEQ ID NO:55:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2162 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:55:

CTAAAAGTGA	AGCCCGATAG	CGTCTCTCTC	CTGCAAGGAT	TTCATAACCA	ATAACAGGAG	60
ATTGACGAAC	AATAATCGGT	TGAATGACCC	CATTTTCTTT	GATAGACTGT	GCTAGTTCAT	120
CTAGCTTTTC	TCTATCAAAT	TCTTTTCGGG	GTTGATAGGG	ATTTTTTTGT	ATATCTGTGA	180
TAGAAATCAT	TTCAAATTTT	TCCATGATTC	TACACTAACA	CATCTTTTCT	CTTATGTAAA	240
GCTTTCCTTA	CATAGATGTC	AATTAAGATT	CTAAATCACC	TGAACTCTTG	TTAAGTTTGA	300
TAGAGGTAGT	TTCTTCTTTC	CCGTTACGAT	AGTAGGTTAT	CTTAATGGTG	TCTCCGATAG	360
AATGGTTGTA	AAGAGCACTT	TGTAAGTCTG	TTGATGAAGC	AATCTCTTTG	TCATCTACTT	420
TTGTAATTAC	ATCGTATTTT	TCAAGGTGAC	CATTGGCAGG	CATATTACTT	TGTACCGAAC	480
GAACAATTAC	ACCAGATGTA	ACATTACTTG	GAATATTGAG	TCTTCTGATG	TCGCTGTGAC	540
TCACATTAGA	TAAATTAACC	ATCTGGATTC	CCAAAGCTGG	ACGCGTCACT	TTTCCGTTTT	600
TTTCTAACTG	TTCAATAATA	TTGATAGCAT	CATTTGCAGG	AATTGCGAAA	CCAAGACCTT	660
CTACAGATGT	TCCTCCATTT	GTAGCAATTT	TACTTGAGGT	AATTCGGATA	ACCTGCCCTT	720
GAATATTGAT	CAGTGGGCCG	CCAGAGTTAC	CTGGGTTAAT	AGCAGTATCA	GTTTGGATGG	780
CTTTTGTTAGA	AATAGCTTGT	CCATCTTCCG	ATTTTAAGGA	TACATTTCTA	TTGAGACTGG	840
ATACGATACC	TTGAGTGACA	GTATTTGCAT	ATTCAGAACC	TAACGGGCTA	CCGATGGCAA	900
TAGCAGTTTC	TCCTACAGTT	AACTTACTAG	AATCACCAAA	CTCAGCTACT	GTTGTCACTT	960
TTTCTGAAGA	GATTTTCGACG	ACAGCAATAT	CAGAGAAAGT	GTCAGCTCCG	ACAATTTCTC	1020
CAGGTACTTT	AGTCCCATCT	GACAATCGAA	TATCTACTTT	GCTGGCGCCA	TTTATAACGT	1080
GATTGTTGGT	GACGATGTAA	GCTTCTTTAT	CATTCTTTTT	ATAAATAACT	CCAGATCCTT	1140
CACTAGAGAT	TCGCTGAGAA	TCTGTGTCAG	TATCATCATT	GCCAAATACG	CTATTTTGTC	1200
TGTTTGCCGA	ATAAGTAATA	ACAGAAACAA	CAGCATCTTT	TACTTTGTTA	ACGGCCTGTG	1260
TTGTTGAATT	TTCCGTTCCCT	TATAGGCAGT	TTGTGTAATA	GTACTATTGT	TGTTAGAGTT	1320
GTTTACACTA	CTTTTTTTGAG	TTAGTTGAGT	TATTGAAAAA	CTACCCAAGG	CTCCACTAAA	1380
AAAGCTAATG	ACGATAACGA	CTAATAATTG	AAACCATTTT	TTGTAAAATG	TTTTTTAGATG	1440
TTTCATATTT	GCCTCCATAT	GTTTGAATTA	CTGAAAGTAT	AAACTGACTA	GCTTAATTAT	1500
AACTTAAACA	CAAAAGTTTT	ACACAAACTG	TGGATAAATC	TTTTTGAACT	GTGATTTTCT	1560
TAATTGAAAT	CTATTTTTTTA	TTTTGTGAAT	AAGATGTGAA	AAAATAGAGA	ATATGTTAGA	1620
ATAGAGTCAT	GAAAATTAAA	GTTGTAACAG	TTGGGAAACT	GAAAGAAAAG	TATTTAAAAG	1680
ATGGTATCGC	AGAGTATTCA	AAACGAATTT	CTAGATTTGC	TAAGTTTGAA	ATGATTGAGT	1740
TATCAGATGA	AAAAACACCA	GATAAGGCCA	GTGAATCAGA	AAATCAAAAG	ATTTTAGAAA	1800
TAGAAGGTCA	GAGAATTTTA	TCAAAAATTG	CTGACCGTGA	TTTCGTTATT	GTGTTAGCCA	1860
TTGAAGGGAA	AAC'TTCTTC	TCAGAAGAAT	TTAGTAAGCA	GTGAGAAGAA	ACTTCTATAA	1920
GGAAGGATGT	CTACTCTTAC	TTTTATTATT	GGGGGAAGTT	TAGGATTGTC	ATCATCTGTA	1980

AAAAATAGAG	CCAATCTTTC	TGTCAGTTTT	GGTCGCCTAA	CCTTGCCTCA	TCAGTTAATG	2040
AGACTAGTTC	TTGTTGAACA	AATCTATCGC	GCTTTTACGA	TTCAGCAGGG	ATTCCCCTAC	2100
CATAAATAGA	GAATTGACTT	TTAATTGAAT	TTTTGGTAGA	ATAATTGTGT	TAGGTCTCAT	2160
AG						2162

(2) INFORMATION FOR SEQ ID NO:56:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1766 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:56:

ATCGAATTTT	CCAAAATGGG	GAGCTAGAGC	AGTGGAGTGA	TTATGTGGCA	GACGATTTGA	60
TTCAGCATAA	TCATGAGATT	GGACAAGGAA	GTGCTGCTTA	TAAAAACTAT	GTGGCTGAAT	120
ATATTGTCAC	TTTTGACTTC	GTTTTCCAAC	TCTTAGGACA	AGGAAACTAT	GTGGTTAGCT	180
ATGGTCAGAC	TCAGATTGAT	GGCGTTGCTT	ATGCCAAGTA	CGATATCTTC	CGTTTAAAGA	240
ACGGGAAAAT	TGTGGAGCAT	TGGGATAATA	AGGAAGTCAT	GCCTAAGGTA	GAAGACTTGA	300
CCAATCGAGG	GAAGTTTTAA	ATTGAGGACA	AAGAATGATT	GAATACAAAA	ATGTAGCACT	360
GCGCTACACA	GAAAAGGATG	TCTTGAGAGA	TGTCAACTTA	CAGATTGAGG	ATGGGGAATT	420
TATGGTTTTA	GTAGGGCCTT	CTGGGTCAGG	TAAGACGACC	ATGCTCAAGA	TGATTAACCG	480
TCTTTTGGAA	CCAATGATG	GAAATATTTA	TATGGATGGG	AAGCGCATCA	AAGACTATGA	540
TGAGCGTGAA	CTTCGTCTTT	CTACTGGTTA	TGTTTTACAG	GCTATTGCTC	TTTTTCCAAA	600
TCTAACAGTT	GCGGAAAATA	TTGCTCTCAT	TCCTGAAATG	AAGGGGTGGA	GCAAGGAAGA	660
AATTACGAAG	AAAACAGAAG	AGCTTTTGGC	TAAGGTTGGT	TTACCAGTAG	CCGAGTATGG	720
GCATCGCTTA	CCTAGTGAAT	TATCTGGTGG	AGAACAGCAA	CGGGTCGGTA	TTGTCCGAGC	780
TATGATTGGT	CAGCCCCAAG	TTTTCCTCAT	GGATGAACCC	TTTTCGGCCT	TGGATGCTAT	840
TTCGAGAAAA	CAGTTGCAGG	TTCTGACAAA	AGAATTGCAT	AAAGAGTTTG	GGATGACAAC	900
GATTTTTTGTA	ACCCATGATA	CGGATGAAGC	CTTGAAGTTG	GCGGACCGTA	TTGCTGTCTT	960
GCAGGATGGA	GAAATTCGCC	AGGTAGCGAA	TCCCGAGACA	ATTTTAAAAG	TGCCTGCAAC	1020
AGACTTTGTA	GCAGACTTGT	TTGGAGGTAG	TGTTTCATGAC	TAATTTAATT	GCAACTTTTC	1080
AGGATCGTTT	TAGTGATTGG	TTGACAGCTA	CAATGACATT	GGTCGGTTCC	TTGAGCAAGA	1140
GATAGATTAG	CCAGACAGTC	ATGCCCAAAA	TCCCTCCAGG	TAAGAGCATA	GACCGTTGCA	1200
CATTAAGTAC	GATTAAAAAA	GTGATAATGG	CAAGAAAAC	TGCTACTGCT	TGTAATAAAA	1260
AGGTTGTTAG	TGTCATATTA	GTTTCATCAAT	ACCAAGGCGA	CAGAAGTTCC	TGCCCCTAAA	1320
GCGAGGGTAA	TGAGCAGGGA	TTCAAACATC	TTACTCATAC	CAGAGTTTAT	GTGGTTGGTC	1380
ATAATATCAC	GGACCGCATT	GGTCAAGGCA	ATACCTGGTA	CAAACGGCAT	GACCGCACCA	1440
GCTATAATCA	AATCTGCCGT	TGAAGGAAAA	CCTGTGTAGC	GAGCCCCAAA	CTGGGCAATT	1500
ATCCCCAAGA	CAAAAGCTCC	AGCAAAGGCT	GTCACAAAGG	GAATTCGGAT	AAATTTTTTC	1560
ACATAGAGGG	AAAAGGCAAA	ACCAAATAAG	GTCGCCACTC	CTGCCCCAAG	TGCGTCGTAG	1620

ATATTTCCGC	TAAACATAAC	TGAAAAGAAA	GGAGCACTAA	AGGTCGCAGC	CAGAGTTACC	1680
TGCAACTTAG	TATAGGGAAG	GGGTTGAGCT	TGCAAGGCCG	TCAATTGCTT	AAAGGCTGTT	1740
TCTAAGTCAA	TCTGCCCCC	AACTGG				1766

(2) INFORMATION FOR SEQ ID NO:57:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1705 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:57:

CTCTGACGGA	GGCTGGTTAT	GTGGGTGAGG	ATGTGGAAAA	TATACTCCTC	AAACTCTTGC	60
AGGTTGCTGA	CTTTAACATC	GAACGTGCAG	AGCGTGGCAT	TATCTATGTG	GATGAAATTG	120
ACAAGATTGC	CAAGAAGAGT	GAGAATGTGT	CTATCACACG	TGATGTTTCT	GGTGAAGGGG	180
TGCAACAAGC	CCTTCTCAAG	ATTATTGAGG	GAAGTGTTC	TAGCGTACCG	CCTCAAGGTG	240
GACGCAAACA	TCCACAACAA	GAGATGATTC	AAGTGGATAC	AAAAAATATC	CTCTTCATCG	300
TGGGTGGTGC	TTTTGATGGT	ATTGAAGAAA	TTGTCAAACA	ACGTCTGGGT	GAAAAAGTCA	360
TCGGATTTGG	TCAAAACAAT	AAGGCGATTG	ACGAAAACAG	CTCATACATG	CAAGAAATCA	420
TCGCTGAAGA	CATTCAAAAA	TTTGGTATTA	TCCCTGAGTT	GATTGGACGC	TTGCCTGTTT	480
TTGCGGCTCT	TGAGCAATTG	ACCGTTGATG	ACTTGGTTCG	CATCTTGAAA	GAGCCAAGAA	540
ATGCCTTGGT	GAAACAATAC	CAAACCTTGC	TTTCTTATGA	TGATGTTGAG	TTGGAATTTG	600
ACGACGAAGC	CCTTCAAGAG	ATTGCTAATA	AAGCAATCGA	ACGGAAGACA	GGGGCGCGTG	660
GACTTCGCTC	CATCATCGAA	GAAACCATGC	TAGATGTTAT	GTTTGAGGTG	CCGAGTCAGG	720
AAAATGTGAA	ATTGGTTTCG	ATCACTAAAG	AAACTGTCTG	TGGAACGGAT	AAACCGATCC	780
TAGAAACAGC	CTAGAGGTGA	CTATGGAAC	TAATACACAC	AATGCTGAAA	TCTTGCTCAG	840
TGCAGCTAAT	AAGTCCCCT	ATCCGCAGGA	TGAAGTCCCA	GAGATTGCCC	TAGCAGGGCG	900
TTCAAATGTT	GGTAAATCCA	GCTTTATCAA	CACTATGTTG	AACCGTAAGA	ATCTCGCTCG	960
TACATCAGGA	AAACCTGGTA	AAACCCAGCT	CCTGAACTTT	TTTAACATTG	ATGACAAGAT	1020
GCGCTTTGTG	GATGTGCCTG	GTTATGGCTA	TGCTCGTGTT	TCTAAAAAGG	AACGTGAAAA	1080
GTGGGGGTGC	ATGATTGAGG	AGTAATTTAA	CGACTCGGGA	AAATCTCCGT	GCGGTTGTCA	1140
GTCTAGTTGA	CCTTCGTCAT	GACCCGTCAG	CAGATGATGT	GCAGATGTAC	GAATTTCTCA	1200
AGTATTATGA	GATTCCAGTC	ATCATTTGTG	CGACCAAGGC	GGACAAGATT	CCTCGTG GTA	1260
AATGGAACAA	GCATGAATCA	GCAATCAAAA	AGAAATTAAA	CTTTGACCCA	AGTGACGATT	1320
TCATCCTCTT	TTCATCTGTC	AGCAAGGCAG	GGATGGATGA	GGCTTGGGAT	GCAATCTTAG	1380
AAAAATTGTG	AGGAAAAGAA	AATGGCAAAA	ACAATTCATA	CAGATAAGGC	CCCAAAGGCT	1440
ATCGGGCCCT	ATGTTCAAGG	AAAAATCGTT	GGCAACCTTT	TGTTTGCTAG	CGGTCAAGTT	1500
CCCCTATCCC	CTGAAACTGG	GGAAATTGTA	GGAGAGAATA	TCCAAGAACA	GACAGAGCAA	1560
GTCTTGAAAA	ACATCGGTGC	TATTTTGGCA	GAAGCAGGAA	CAGACTTTGA	CCATGTTGTC	1620
AAAACAACCTT	GTTTCTTGAG	CGATATGAAC	GACTTTGTTC	CTTTTAATGA	GGTTTACCAA	1680

ACGGCCTTCA AAGAGGAATT CCCAG

1705

(2) INFORMATION FOR SEQ ID NO:58:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1673 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:58:

ACGTTTTGGG	AACTGTTCCG	ATAGCAGATT	CCGAACAAAC	TGATAATGGT	TGGCAAAATC	60
ATTATTCCTA	ATAGTAACGA	AGCTGGTTAG	GACAACTCAT	GCCATTTTCCT	AAAAAGGTTT	120
TAATCCAAGG	CACCAATAAT	TGTAGGCCGA	AAAAACCATA	AACAATAGAT	GGAATGGCTG	180
CCATCAAGTT	GATAGCTGAT	TTTAAGAAGC	TATAGACGGG	CTTTGGACAA	TTATAAACCA	240
TAAACACCGA	TGTCAAGATC	GCCTGTTGGC	ACCCCAATCA	CAATCGCTCC	TAAGGTCGAA	300
TAAATAAGGA	ACCAACGATC	ATTGGTAAAA	TACCATAGCT	TGCCGGAATG	TTCTGTTGGCG	360
ACCAATCACT	GCCTAATAAA	AAACGGGCAA	AGCCGTAGTT	AGCTATGAAA	GGTAAGCCAT	420
TACTAAAAAT	AAAGAAACAG	ATTAGCAAAA	TAGCTACAAC	AGCTACTGTT	GCACTCATGA	480
AAAAAATTGC	CCTAAAAACT	GCTTCTTTGA	AGGCTTGTTT	TGTCACATCT	TGTCCTTTCT	540
AGTGAAGAAA	GTAAGGGAGA	TACGACACCT	CCCTACTTGC	CTTCTTTATC	TTATTGTACG	600
ATGAAACGTC	TGCATCTCTT	TAGAGATTTA	TGGAGCAAAC	ATTTTATTTA	ATCTTGTCCC	660
AGGTGGTTAA	TTTGCCACTA	AAAACGTCCG	CAAGTTCAGC	CATACTGACT	TGGCTTGCCT	720
TATTGTCATT	ATTGACCACA	ACAGCAATAC	CGTCTAAAGC	AATAGCATCA	TGGGTGAGAC	780
TCTTACCTTC	TTCAGGAGTT	AATTCCCTAG	AAACCATAAC	AATATCAGCG	GTTTTCTCCT	840
TAACAGCGGT	AATACCTGCT	GAAGACCCAT	TAGAGGTAAT	ATCAATCGTA	ACTTCTGGAT	900
TTTCTTTTTT	ATAAGCTTCT	GCTAATTTTT	CCATTAAAGA	AGATACTGAA	GTGGAACCTA	960
CAACAGACAA	CTTGCTGAT	AAGTGTTGGC	TTGTATATTC	TGTGGTTTCG	GTTTTAGCTT	1020
CAATAAATTT	ATTATCTGTG	ACCACTTGTT	GACCTTGTTT	GGAGTGGATA	AAGCTGATAA	1080
AATCTTGACC	TAGCTTGGA	AGATTAGAAG	ACCAAACAAT	GTTGAAGGGA	CGTTGAAGAG	1140
GGTATTCACC	ATCTAAAAC	GTGTCTCGAC	TAGCCTTGAC	ACCATCAATC	TCTAAAGCCT	1200
TGACAGATTT	CGTTAAAGAT	CCCAAGGAGA	TGTAGCCGAT	AGCATTAGCA	TTCCCTTGAA	1260
CTGCTGAGAG	AACACCTTCT	GTACTATTTT	GAATCACAGC	TGTTTTGGCA	GTGTAGTCAA	1320
TTTTTTTATC	ACCGTCTTTT	TTGAGAAATC	CTGTGATTTT	TGTGAAGGCA	CCCCGTGTTT	1380
CAGAGCCATT	TTCTCGTGAA	ATCACCTCAA	TCGTTCCCTG	AGCTGACTGT	TTGGAAGCAG	1440
CTGACTGATT	GCCACAGGCA	ACAAGCCCCA	ATCCTGATAA	GCCAATGGCT	GCAAGAGTAA	1500
GCATTTTTTT	GAATTTTCATA	ATAATCACCT	TTATCTCTAT	GTATTTTTTCT	TGTGTAGGCT	1560
TACTACATTT	ATAGTCTAAC	AAGTCTTTGT	AAAGGTTTAT	CCCTGATTCA	TGTAAAGATT	1620
GTGTAAAGAA	TCAAAAAAAG	CCACTTTTGA	AAAATGGCTG	CCCCTAAAAA	TAG	1673

(2) INFORMATION FOR SEQ ID NO:59:

172

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1702 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:59:

```

CTTTTTTATT TCACAACAAG TTCATAACGT GTCTTACTGG TGAAGGTTTG ACCAGCTTTA      60
AGAATGACTT GGCCTTTAAG GTCACTGTGA ATGGCATCTG GTAAAGCTTG CGCTTCAAGA      120
GCAATCCCAT TGTGCTGTAG CATTGGCTGA CCTCCTATGA TGACACTTTC ATCCACAAAG      180
TTTGCTGTGT AGACCACAAA GCAAGGAGCT TCTGTCTTGA AAAGCAGGAA GCGACCTGAA      240
TTTTGGTCAT AAAGGAATCC AGCATTGTCA TGGCCTGCAG GAAGGGCAAA TGGATGATCC      300
AAACCTGATG CCAGCTGGAT TTGCTCATCT TCTTCTGCAA AGATATCCTT CAACAAGGCA      360
CCATTGTAGA TGTGTTTGAC CACATCACGG TTGGCTTCTG GAGTTTGGC AGGAACACCG      420
TCAGGAGCGA TTGAGTAAAT GCCCTCTGTG TTTAGTTGGA AGACATGACG GTCAATCGTC      480
TGCGTGAAAT CACCAGACAA GTTGAAATAG CTGTGGTTGG TTGGATTGAC CAGCGTATCC      540
TGATCGGTCG TTACCTTGTA GATCGAATTC ATGGAGGCAC CAGTTTCTTC CAAGTGATAA      600
CTGATCGCCA AATCTTGAGA TTTCCAGGGA ACCCTCCTGT CCCATCTGTA CGCTCTGTGT      660
AGAGAGTCAA GCCATGATCG CTTACTTCTT CAACTTCAAA CAAGCTGGAA TCCCAACCAG      720
TTGAACCACT GTGATTACAG TTGCTAGCAT TATTAACCTC AAGGTCATAG GTCTTACCAT      780
TGAGCTCAAA GGTGCGACCT GCAATACGAC CCGCTACAGG ACCTACACTT GCTCCATGCT      840
TGGGACTATT GCCTACATAA CTATCAAAGT CATCAAATCC CAAGATAACA TTGGCAAAAT      900
TTCCAGCCTT GTCAGGTGCG ACATAGCGCA AGATAGTCGC ACCATAAGTC ATAACCTCAA      960
GTTGGTAGCC ACCGTCTGTC TCAAATCGAT AGGCCAAGAC ATCCTCACCC TCAACATTTT     1020
CAAATACACG CTCTGTGTAT GCTTTCATTC TGTTCTCCTT TTACTATTTT TCTCAAGCAA     1080
ACAAACCATA GAAAGCGTAC TGACAATCTA TGGTTTATCT GATAATTTAC AAATCCTCTT     1140
GTCAAGAATT CATAAACACT GTCTTACTTT TGATATTCGT GAATTATGAC ACCTTGTA     1200
ACACGGTTTA CTGTACCTGT AGGAGACGGT GTATCTGGTT TATTTTCTAC CTTGAGTGAA     1260
GTCAATAGGG CAAAGAGTTG GGCATAAACG ATGTAAGGGA AGACACGGTA AATATCATTC     1320
AAGACACCGC CACAACCAAG GGCCACTTCT TTGACATTTT CAAGACCAAA AGCTTGATCA     1380
CTCAAAAGCA CAACACGACG AGCAATCTGG TCACCAGCAA CTTACGAAAC CAAGTCCAAG     1440
TCGTACTTAC GAGTGTAGTC CGTCGTTGTA CCAAAGACCA AAACAACGTG ATTGTCGTTG     1500
ATAAGAGATT TTGGACCGTG ACGGAAGCCA ACTGGGCTTT CATACATGGT CGCAACTTGA     1560
CCAGCAGTTA ATTCCAAAAT CTTGAGCTGA GCTTCATGAG CAAGTCCAAA GAAAGGACCA     1620
GCGCCTAGAA TAGATGACAC GGTAAAGTC TAAATCAACG AGATCTTTGA CATCTTCTGC     1680
CTTGTCTAAA ACTTTACGGG CA                                     1702

```

(2) INFORMATION FOR SEQ ID NO:60:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1940 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:60:

TGCAGGATTT	GATTTGGACG	ACTTTTATTA	TTACCAGATT	CGCCTAGGAA	TAGAAAAAAG	60
AGCCCAAGAG	TTGGACTATG	ATATCTTGCG	CTATTTTAAT	GACCACCCCT	TTACCCTAAG	120
CGAGGAAGTG	ATTGGGATTC	TCTGCATCGG	AAAGTTTAGT	CGAGCTCAGA	TTTCTGCCTT	180
TGAAGAATAC	CAAAAGCCTC	TTGTATTTCT	AGACAGCGAT	ACACTTTCCT	TGGGACATAC	240
CTGTATTATC	ACGGATTTT	ACACTGCTAT	GAAACAGGTT	GTCGATTATT	TCCTCAGTCA	300
AGGAATGGAC	CGTATCGGGA	TTCTAACAGG	CCTTGAAGAA	ACAACAGACC	AAGAAGAAAT	360
CATTCAGGAC	AAGCGTCTAG	AAAACCTCAA	AAACTACAGT	CAAGCGAGGG	GAATCTATCA	420
TGATGAACTG	GTCTTTCAAG	GAAGATTTAC	TGCCCAGTCT	GGCTATGACT	TAATGAAGGA	480
GGCCATTCAG	AGCTTGGGAG	ACCAACTTCC	GCCAGCATTT	TTCGCAGCCA	GCGATAGTTT	540
AGCTATCGGT	GCCCTCCGTG	CCCTCCAAGA	AGCTGGAATC	AGCCTGCCAG	ATCGCGTCAG	600
CCTCATTTCC	TTTAACGACA	CTAGTCTGAC	CAAACAGGTC	TATCCTCCCC	TCTCTAGTAT	660
TACAGTTTAT	ACTGAAGAAA	TGGGCCGAGC	AGGTATGGAT	ATTCTTAACA	AGGAAGTCCT	720
CCACGGTCGG	AAAATCCCTA	GCCTGACCAT	GCTGGGAACC	AGACTGACAT	TAAGAGAAAG	780
TACCCTAAAT	CAAGAATAGG	ATAACATAAA	AAACGAATAG	AGTTCTAAAA	CTCCTATTCG	840
TTTTTTTATC	GATTACAATC	ATAGACTTAA	TGGTCTTACG	TTCATCCATA	TCTTTGTAGG	900
CTTGGTCGAT	ATCTTCCAGT	TTATAACTTG	AAGTAAAGAC	GCGACCTGGA	TTGATATCAC	960
CATCAAGGAC	GGCTTTTAGT	AAAAATTGCT	TATCGTATGT	TGTAGCAGAA	GCTGCCCCAC	1020
CTGCTACAGA	GATATTTTGC	ATAAATGTCT	AACCAAGAGC	ACGATTATTA	TAGTGTGGGA	1080
CTCCTACAAA	GCCCATACGC	CCTCCATTAT	GAAGAACACC	TAGCGCCTGT	TCTATAGCAG	1140
CCTCCGTACC	AACACATTCA	AGTGCTGCGT	CTGCTCCTCC	GCCGAGGATT	TCACGCACCT	1200
TGGTAATTCC	TTCTTGACCA	CGTTCCTGCA	CAACAGCTGT	CGCACCTGAC	TCCATAGCCA	1260
TCTTTTGACG	GTCTTCATGA	CGGCTCATAA	GGATAATTTG	TGATGCTCCA	CGCATCTTAG	1320
CCGCGATGAC	AGCACATTGA	CCAACAGCCC	CATCACCGAT	AACAACAACC	TTGTCCCCTT	1380
TTTGAACATT	TGCAACACGC	GCCGCATGAT	AGCCTGTCTG	CATGACATCT	GCAAGAGTCA	1440
AAAGGGACTT	GAGCATCCCT	TCTGTATAGT	CAGAAGGTTG	ACCAGGGATT	TTAACCAGCG	1500
CCCAGTTTGC	ATAGTGGAAG	CGAATATATT	CTGCCTGAAA	ATCACCCCCC	AAATTATTGC	1560
CAATATGATT	GTCGCAAGAA	CCGTCAAATC	CAGCAAGACA	GGCATCACAC	TCACCACATC	1620
CATGTGTAAA	AGGGACAATC	ACAAAATCAC	CTGGTTTCAC	CGTCGTAATG	GCTTCCCCAG	1680
CTTCTTCAAC	AATCCCAATC	GCTTCGTGTC	CACTTATTTT	TTGTGTCCAA	CTTTCGTTTT	1740
CCNTGGATTA	CGGTACCTCC	ATAAATTTGA	ACCACAAACG	CACGCACGAA	CCACACGAAT	1800
AATCACATCA	TCCGCTTCTA	TTATTTGCGG	ACGTTCAATG	CTAGCAAGTC	CAACCTGACC	1860
TGCCTTTGTA	TATACTGCTG	ATTTTCATTTA	AAATTTTCCT	TCCTTATAAA	GTTTAATTTT	1920
GAGATTTAAA	CGATTTAAAG					1940

(2) INFORMATION FOR SEQ ID NO:61:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2051 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:61:

ATCGAATTTT	TCTAGCCAGG	CTACAGTTTT	GGCAAGTAAG	GTTTCATCTC	AGGCAGTCAA	60
CTGGGTGAGT	GCCTTTATTA	GCGGAGCTTC	TCAAGTGATT	GTTGCCTTGA	TTATCGTTCC	120
TTTCATGCTC	TTTTATCTCT	TGCGTGATGG	GAAAGGCTTG	CGTAACTATT	TGACCCCAATT	180
CATTCCAAGA	AAATTGAAGG	AACCTGTTGG	ACAAGTTCTA	TCAGATGTGA	ATCAACAGTT	240
GTCCAACAT	GTTTCGAGGC	AAGTGACAGT	GGCTATTATT	GTAGCAGTAA	TGTTTTATCAT	300
CTTCTTCAAG	ATTATTGGTC	TACGCTATGC	GGTTACGCTG	GGGGTTACTG	CTGGTATTTT	360
AAATCTGGTC	CCTTATCTTG	GTAGCTTTC	AGCCATGCTT	CCTGCCCTAG	TATTGGGTTT	420
GATTGCTGGT	CCAGTCATGC	TTTTGAAAGT	AGTGATTGTC	TTTATTGTAG	AACAACTAT	480
TGAAGGCCGT	TTTGTCTCTC	CATTGATTTT	GGGAAGTCAA	TTAAACATCC	ACCCTATTAA	540
TGTTCTCTTT	GTTTTGTAA	CTTCAGGATC	TATGTTTGGT	ATCTGGGGAG	TTTTACTTGG	600
TATTCCGGTT	TATGCCTCTG	CTAAGGTTGT	CATTTTCAGCC	ATTTTCGAAT	GGTATAAGGT	660
AGTCAGTGGT	CTATATGAAT	TAGAGGGTGA	GGAAGTCAAG	AGTGAACAAT	AGTCAACAGA	720
TGTTACAGGC	TTTGAGGAG	CAAGATTTAA	CTAAGGCTGA	GCATTATTTT	GCCAAAGCTT	780
TAGAAAATGA	TTCAAGTGAT	CTTCTGTATG	AGTTGGCAAC	TTATCTTGAA	GGGATTGGTT	840
TCTATCCTCA	GGCCAAGGAA	ATTTACCTGA	AAATTGTAGA	AGAATTTCCA	GAGGTTTCATC	900
TTAATCTAGC	TGCAATGGCT	AGCGAGGATG	GTCAAATAGA	AAAAGCCTTT	AACTATCTTG	960
AGGAAATCCA	AGCTGACAGT	GACTGGTATG	TCTCGCTCTT	TGGCTCTGAA	GGCAGACCTA	1020
TACCAGCTGG	AAGGTTTGAC	AGATGTGGCA	CGTGAGAAAT	TATTGGAGGC	CTTGACCTAC	1080
TCAAAGGATT	CTCTCTTGAT	ATTGGGTTTG	GCAAAGTTGG	ATAGTGAGTT	GGAAAAATTAC	1140
CAAGCGGCTA	TTCAAGCCTA	TGCCCAGTTA	GATAATCGCT	CGATTTATGA	GCAAACGGGC	1200
ATTTCCACCT	ATCAACGAAT	TGGCTTTGCC	TATGCTCAGT	TAGGGAAATT	TGAAACGGCT	1260
ACTGAGTTTT	TAGAAAAAGC	CCTGGAGTTA	GAATACGATG	ACTTAACAGC	TTTTGAGTTG	1320
GCCAGTCTTT	ATTTTGATCA	AGAAGAATAT	CAAAAAGCCA	CCCTCTACTT	TAAGCAGCTT	1380
GATACCATTT	CTCCTGACTT	TGAAGGCTAT	GAGTATGGGT	ACAGTCAGGC	TTTACATAAG	1440
GAACATCAAG	TTCAAGAAGC	CCTGCGTATC	GCTAAGCAAG	GATTAGAGAA	AAATCCCTTT	1500
GAAACTCGCC	TCTTGCTAGC	TGCTTCACAA	TTTTCTTATG	AATTGCATGA	TGCTAGTGGT	1560
GCAGAAAATT	ATCTCCTTAC	TGCAAAAAGAA	GACGCTGAGG	ATACAGAAGA	AATCTTGCTT	1620
CGTTTAGCCA	CTATTTATCT	GGAGCAGGAG	CGTTATGAGG	ATATTCTAGA	CTTGCAGAGT	1680
GAGGAGCCAG	AAAATCTTTT	GACCAAGTGG	ATGATTGCTC	GTTCTTATCA	AGAAATGGAC	1740
GATTTGGATA	CTGCTTATGA	GCATTATCAA	GAGTTGACAG	GAGATTTGAA	GGACAATCCA	1800
GAATTTCTGG	AACACTATAT	CTATCTCTTG	CGTGAATTGG	GACATTTTGA	AGAAGCAAAA	1860
GTCCATGCTC	ACACTTACTT	AAAAGTGGTT	CCAGATGATG	TGCAAAATGCA	AGAACTGTTT	1920

GAGAGATTGT	AAGAATGTTT	AAACATATAG	AACTGTAGTT	TATCTCTTTT	GATAGCTACG	1980
GTCTTTATTT	GTACATGGTA	GAATCTTTTT	ACAAAAATAC	TTGGTAATCT	TGTTTATTCA	2040
TGCCATAATA	G					2051

(2) INFORMATION FOR SEQ ID NO:62:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1318 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:62:

CTTTAGCAAT	CAGTTTATTG	GGAGATTTGA	CTGCCACTTC	TGTTGGAACC	TTGATAATCT	60
TTTTACCCTC	AAAGCGTTCC	ATACCAGAAA	TCTTAACATC	AACTGCTAAA	ATAACTACAT	120
CCGCTGCATC	AATCTGCTCT	TGACTCAATT	CATTTTCTAC	CCCTATTGTC	CCCTGAGTCT	180
CAACATGAAT	CACATGTCCA	GCTACCTTTG	CGGCATTCTC	TAATTTTTC	TGTGCAATAT	240
AAGTGTGGGC	AATTCCCATA	GTACAAGCTG	CAACACCAAC	AATTTTCATA	CGGATACCCT	300
CCAAAATTTT	TTCTTATTAA	CAAAAAGCTG	CAATCACATC	ATCAGATGTC	TGAGCCCGAA	360
CTAATTTGGC	AACAACFTCG	TCATTACCAA	GTTTTTCGAGC	AAAGAGTGAT	AAGGTCTTCA	420
AATGCTCCCT	AGCAGCTTCT	GTATCATCAC	CAACTGCAAA	GAGTACAATT	ACTTTGACCC	480
CTTTCCCATC	AATGGTCTCC	CAAGGAATCT	CATTGTGATT	TATAGCTATG	ACTACCCCGG	540
CCTTCTCCAC	AGCAGAACTC	TAGCTATGGG	GAATAGCAAT	ATAATTCCCA	ATACCGGTCT	600
GTCCTTCTGC	CTCTCTCTGA	TAAAGACCTT	CGATAAATTG	GTCTCTATCA	GACACATAAC	660
CCGTCTCAAC	CAATAGTATG	AGCTAATGCC	TCAAAAACCT	CTTCTTTGCT	CTGCATCTGT	720
AAATCCGTCT	GGATCAGACT	CACATTAAGA	ATATCTTTGA	TTTCCATATA	TTATCTCCCG	780
TAATTCCTTCT	TTTGTTAACT	GTTTTAATTG	ATTTATGAAT	GATTCATCTG	CTAGTCTTCT	840
CATCAATGTT	TTAATACATG	ACTTGTCCTG	TGATACTGCA	ATGGCCAAAC	CGATAATAAG	900
GTCAACACAC	TGGATATCCT	TCGACCATTG	TCTGATAGGT	GGTTTAAATC	TAGTAATCAC	960
TAAGACATGA	TGTTGAAAGT	TTCCTTCACA	ATGTGGTAGA	AGAACACCTT	TAGCAACCTC	1020
TATACTTCCC	TGTCTCTCAC	GGTAATATAG	AAGCTCTTCT	ATTTTTTCTG	TATCTTCAGA	1080
AACAAGAAGG	CTGATTTGAT	TTGCTAATTC	TTTGTAGGCT	TCTTGACGAT	TTTGAACAGA	1140
TATATCCATA	AGGACAAGCG	AAAGATTATT	CATAGTTTAT	CTCCTGAATT	TTTGCTTGAA	1200
GACGTTGTTT	ATCACCCCTG	GTTAGAAAAG	CACTAACTAG	GACAAACGGG	ACACTTGCTG	1260
GTTCTTGCAA	AGCTACCGTC	GTCACAATGA	AATCTAAATC	TGGATATAGA	TTTATCAG	1318

(2) INFORMATION FOR SEQ ID NO:63:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2077 base pairs
- (B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:63:

CTAGTCTTGG	CTACTGTCTA	AGTTGGCTTG	TGCATAAGCC	TGCCAGATTT	TTTGTGTTGGG	60
TTTGGCAAGT	GGGTAATTCT	TGAATTCTTC	TGGTGAAAGC	CAACGAACTT	CCCTATCTGA	120
AAAATCATGG	AAGTCACTCA	CCTGACCTGC	TACAATCTGT	ACATGCCATT	TTCGATGACT	180
AAAAACATGC	TGGACTGTAT	CAAAACAAAC	ATCAAGCCAA	TCAACATCTA	GGTCATAGTC	240
CTGCTGGAAA	CTCTCTTCTG	GGACTGGGGC	CAGAGTTCAC	ACTTTCTTCC	GCAACCTGAT	300
GAAAGAGGTC	AAACTGCTCT	TCTTGCGAAA	AGTTATCAAC	TTCTATAAAG	GGGAAATGCC	360
AAAAACCTGC	CAAGAGCTTT	TCGCTTTCAT	TTTTTTCAAG	TAAAAATTGT	CCTTGAGAAT	420
TTTTCACAAC	TAAGGCTTTA	AGATAAATAG	GAACCGGCTT	TTTCTTAGGA	GATTTAATTG	480
GATAACGGTC	CATGGTTCCA	TTCTGATATG	CCGCACATAA	GTCCTTGACT	GGGCTTTCTT	540
CAGGTCGGG	ATTTACAGGA	GACTCAATAT	CAGACCCTAA	GTCCATCAAG	GCTTGATTAA	600
AATCACCCGG	ACGATCTGGA	TTAATCAAGA	TCTCCATCAT	TGCCTGAAAA	ATTTTTTCGAT	660
TACTTGGAAT	CCCAATATCG	TGGTTGACTT	CAAACAGACG	CGCCAAGACC	CGCATGACAT	720
TACCATCTAC	AGCTGGCTCA	GGCAAAGTTA	AAGCAATACT	GGAAATGGCT	CCTGCTGTGT	780
AAGGTCCAAT	CCCTTTCAAG	CTGGAAATTC	CTTCATAGGT	ATTTGGAAAT	TGGCCACCAA	840
AGTCAGTCAT	AATCTGCTGG	GCTGCAGCCT	GCATATTGCG	AACTCGAGAA	TAATAACCCA	900
AGCCCTCCCA	AGCTTTCAGT	AAACTCTCCT	CAGGCGCAGT	TGCCAGACTT	TCGACAGTTG	960
GAAACCAGTC	CAAAAATCTT	TCGTAGTAAG	GGATAACTGT	ATCCACCCTG	GTCTGCTGAA	1020
GCATGATTTT	AGATACCCAG	ATGTGATAAG	GATTTTTTACT	TCTCCTCCAA	GGCAAATCTC	1080
TTTTGTTTTC	ATCATACCAA	GCGAGAAAGT	TTCTCACCGG	AAAGAAATGA	CTTTCTCCTC	1140
CGGCCACATG	ACGATACCGT	ATTCTTTCAA	ATCCTAACAT	ATCTCTAGTT	ATAACACAGA	1200
AGGTTTCACC	TGTCTTTGTA	TCTGATTTAT	AATATTTTCA	ATAGATAGTA	TATAACTTTT	1260
CCTATCTACT	TATACTCCAA	TGAAAATCCA	AAGAGCAAAC	TAAGAAGCTA	GCCGCAGGTT	1320
GCTCAAAACA	CTGTTTTGAG	GTTGTGGATA	GAAC TGACAG	AGTCAGTATC	ATATTACCTA	1380
CGGCAAGGTG	AAGCTGACGT	AGTTTGAAAA	GATTTTTCGAA	GAGTATAAAT	CTTATTGATG	1440
AACTGCTTGC	AGTCTGAGAA	AAAATGAGCT	TGGATATTAT	TTCCAAACTC	ACTTAAAGTC	1500
AATTTCAATC	CACTAGAACA	AGCCTAGTAC	AGTTCCATCG	CTTTCAACAT	CCATGTTGAG	1560
AGCTGCTGGA	CGTTTTTGAA	GACCTGGCAT	GGTCATAACA	TCACCAGTTA	AGGCAACGAT	1620
GAAGCCTGCA	CCTAATTTTG	GTACCAATTC	ACGAATGGTA	ATTTCAAAGT	TTTCTGGTGC	1680
TCCAAGCGCA	TTTGGAATTGT	CTGAGAAACT	GTATTGAGTT	TTAGCCATAC	AAATTGGCAA	1740
TTTGTTCCCA	CCGTTTTGAA	CGATTTGAGC	AATTTGTGTT	TGAGCTTTCT	TCTCAAAGTT	1800
CACTTTGCTA	CCACGATAGA	TTTCAGTGAC	AATTTTTTCA	ATCTTTTCTT	GGACAGAAAG	1860
GTCAATTATCG	TACAAACGTT	TATAGTTAGC	TGGATTTTCA	GCAATTGTCT	TAACAACTGT	1920
TTCTGGCAAGT	GCTACTCCAC	CTTCTGCTCC	ATCAGCCAG	ACACTAGCCA	ATTCAACTGG	1980
TACATCGATT	GAGGCACAGA	GTTCTTTTAA	GGCTGCAATT	TCAGCTTCTG	TATCAGATAC	2040
AAATTCGTTA	ATAGATACAA	GCTAATGGAA	TACCGAA			2077

(2) INFORMATION FOR SEQ ID NO:64:

177

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1887 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:64:

```

CTCAAAACNC TGCTTTGAAG AGATTTTCAA AGAGTACAAG AAGTTTAGTT ATTAGCGTTC      60
TTACCGCTTG TAAACTAGAT TTCTCATAAA ATAGAATCTT TTCCTTTTAG TTGTAAACTA      120
GTCTGGGAGA GTAGAGAGGT TTGAGATACC TTTCTAGCTT TTGGATTATC ATCTAAGAAG      180
AGTAATTTCC CTTGCATTAA AAAGGGGAAA AAGAGACACG AAATGACTAT AATGGGTGAC      240
AATGGGGGAA GGGATAGACA AGAGATTTTA TCCACATATG AAAAAAGGAG GTTAGGAAAG      300
AGTTATATAT CCTATATTAT ATAAATAATC AATTGCGCAG AAATTTGGTA AGAATTCATG      360
CGTCAACTCA TAAAGAACTA CTTAAAAAAT TCACAGTATT CATAATTATT TTCGAGGAGA      420
AAAACAGTGA AAAAAAGAAA AAAGCTTGCT CTGTCTCTTA TCGCTTTTTG GCTGACGGCT      480
TGTTTAGTAG GCTGTGCTAG CTGGATTGAT CGTGAGGAAT CCATAACGGC TGTTGGCTCA      540
ACTGCCTTGC AACCCTTGCT TGAAGTAGCG GCAGATGAAT TTGGCACCAT CCATGTTGGA      600
AAAACGGTCA ATGTCCAAGG GGGAAGTTCT GGTACAGGCT TGTCCCAGGT TCAGTCTGGG      660
GCAGTTGATA TAGGAAACTC AGATGTATTT GCTGAGGAAA AAGACGGAAT TGATGCTTCT      720
GCTCTTGTTG ACCACAAGGT CGCGGTAGCT GGCTTGCTC TGATTGTCAA TAAGGAGGTT      780
GATGTTGATA ACCTAACGAC AGAGCAACTT CGTCAAATCT TCATAGGTGA GGTAACCAAT      840
TGGAAAGAGG TTGGTGGTAA GGACTTACCC ATCTCTGTTA TCAATCGGGC AGCCGGCTCT      900
GGCTCTCGTG CTACCTTTGA TACTGTCATT ATGGAAGGTC AGTCTGCCAT GCAAAGTCAG      960
GAGCAGGATT CAAATGGAGC GGTAATAATCA ATCGTATCAA AAAGTCCAGG AGCTATCTCT     1020
TATTTATCTC TTACCTATAT AGATGATTCT GTCAAAAGCA TGAAGTTGAA TGGCTATGAC     1080
TTAAGTCCAG AAAATATAAG TAGCAATAAT TGGCCCTTGT GGTCTTATGA GCATATGTAT     1140
ACATTGGGGC AGCCCAATGA GTTGGCTGCA GAATTTCTCA ATTTTGTTCT CTCGGATGAG     1200
ACCCAAGAAG GGATTGTCAA AGGATTGAAG TATATTCCGA TTAAGGAAAT GAAGGTTGAA     1260
AAAGATGCTG CCGGAACGTG GACAGTGTTG GAAGGGAGAC AATAATGAAT CAAGAAGAAT     1320
TAGCTAAGAA AATGTTGCTT CCATCAAAGA ATTCTCGTCT GGAGAAATTA GGAAAAGGTT     1380
TGACCTTTGC CTGTCTTTCT TTGATAGTCA TCCTTGTTGG CATGATTTTG GTTTTCGTAG     1440
CGCAAAAAGG CTTGTCGACC TTCTTTGTCA ATGGTGTGAA TATCTTTGAC TTTCTTTTGG     1500
GAGGAACTTG GAATCCTTCT AGTAAAGAAT TTGGTGCCCT TCCTATGATT TTGGGTTCCCT     1560
TTATCGTTAC CATCTCTCA GCCCTTATCG CAACACCCTT TGCTATTGGT GCAGCAGTTT     1620
TTATGACCGA AGTATCACCA AAAGGGGCGA AGATTTTGCA ACCAGCTATT GAACTCCTGG     1680
TTGGGATTCC TTCAGTAGTG TACGGATTTA TTGGCTTGCA AGTCGTCGTT CCCTTTGTTT     1740
GCAGTGTCTT TGGTGGGACT GGTTTTGGGA TTTTGTGAGG GATTTCGTC CTCTTTGTCA     1800
TGATTTTGCC GACCGTAACC TTTATGACAA CGGATAGCTT GCGTGCGGTT CCTCCNTTAT     1860
TATCGTGAAG CCAGTTTCGC TATGGGA                                     1887

```

(2) INFORMATION FOR SEQ ID NO:65:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 405 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:65:

CTGAGGAATC	AAAAGTTGAA	CCACCAGTAG	AACAAGCATA	AGTCCCAGAA	CAACCCGTGC	60
AACCTACACA	AGCTGAGCAA	CCAAGTACAC	CAAAAGAATC	ATCACAACAA	GAAAATCCTA	120
AAGAAGATAG	GGGAGCGGAA	GAGACTCCGA	AACAAGAAGA	TGAACAGCCA	GCAGAAGCCC	180
AAGAAATCAA	GGTTGAAGAA	CCAGTAGAAT	CTATAGAGGA	GACTGTCATT	CAACCTGTTG	240
AACAACCAAA	AGTGGAACG	CCTGCTGTTT	AATAACTAAC	GGAACCTACA	GAGGAACCTA	300
AAGTTGAAGT	AACTAGTATT	CCCCTACTA	CTCGCTATGA	GGAAGACCTT	ACTTACGAAC	360
ACGGAACGCG	TTGAAGTTGT	TAAGGAAGGT	TATAATTGGC	AGTAT		405

(2) INFORMATION FOR SEQ ID NO:66:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1542 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:66:

CTATGGGATT	GGTAGTCTT	CCTAGTGCAG	GGGCTGTAGA	CCCAGTTGCG	ACCCTAGCGC	60
TGGACTAGTC	GAGAGGGTGT	TGTTGAAAAT	GGATGGCTAT	CGCTATGTTG	GTTATCTATC	120
AGGTGACATC	CTCAAAACGC	TTGGCTTGGA	CACTGTTTTA	GAAGAAACCT	CAGCAAAACC	180
TGGAGAGGTG	ACTGTAGTCG	AAGTTGAGAC	TCCTCAATCA	ACAACAAATC	AGGAGCAAGC	240
TAGGACAGAA	AACCAAGTAG	TAGAGACAGA	GGAAGCTCCA	AAAGAAGAAG	CACCTAAAAC	300
AGAAGAAAGT	CCAAAGGAAG	AACC AAAATC	GGAGGTAAAA	CCTACTGACG	ACACCCTTCC	360
TAAAGTAGAA	GAGGGGAAAG	AAGATTCAGC	AGAACCATCT	CCAGTTGAAG	AAGTAGGTGG	420
AGAAGTTGAG	TCAAAACCAG	AGGAAAAAGT	AGCAGTTAAG	CCAGAAAGTC	AACCATCAGA	480
CAAACCAGCT	GAGGAATCAA	AAGTTGAACC	ACCAGTAGAA	CAAGCAAAAG	TCCCAGAACA	540
ACCCGTGCAA	CCTACACAAG	CTGAGCAACC	AAGTACACCA	AAAGAATCAT	CACAACAAGA	600
AAATCCTAAA	GAAGATAGGG	GAGCGGAAGA	GACACCGAAA	CAAGAAGATG	AACAGCCAGC	660
AGAAGCCCAA	GAAATCAAGG	TTGAAGAACC	AGTAGAATCA	AAAGAGGAGA	CTGTTAATCA	720
ACCTGTTGAA	CAACCAAAAAG	TGGAAACGCC	TGCTGTAGAA	AAACAAACGG	AACCAACAGA	780

GGAACCAAAA	GTTGAAGTAA	CAAGTATTCC	CCAACTACT	CGCTATGAGG	AAGACCTTAC	840
TAAGGAACAC	GGAACGCGTG	AAGTTGTTAA	GGAAGGTAAG	AATGGCAGTA	GAACAGTTAC	900
TACTCCATAT	ATCTTGAATG	CGACAGATGG	TACGACTACA	GAAGGCACTT	CGACAACCTGA	960
TGAAGCTGAG	ATGGAGAAAAG	AGGTTGTTCG	TGTTGGCACG	AAACCCAAAAG	AAAAATTAGC	1020
TCCAGTCTTA	AGTTTGACAA	GTGTTACAGA	TAATGCAATG	TTGCGTAGTG	CGAGACTTAC	1080
TTATCATTTG	GAAAATACAG	ATAGTGTTGA	TGTGAAAAAA	ATTTCATGCTG	AAATTAAAAA	1140
TGGCGATAAG	GTTGTCAAAA	CTATTGACTT	ATCTAAAGAG	AGATTATCAG	ATGCTGTTGA	1200
CGGTCTTGAA	CTTTATAAAG	ATTATAAGAT	TGTGACGAGT	ATGACCTATG	ATAGAGGTAA	1260
TGGTGAAGAA	ACCTCTACGT	TGGAAGAAAC	TCCACTACGA	TTAGACCTCA	AGAAGGTTGA	1320
ATTGAAAAAC	ATCGGCTCTA	CTAATCTCGT	CAAAGTAAAT	GAGGATGGTA	CTGAGGTGGC	1380
AAGTGAAGTTC	TTAACAAGTA	AACCTGTGGA	TGTGCAGAAT	TACTACCTCA	AAGTAACTTC	1440
CCGTGATAAT	AAAGTTGTTT	CCCCTCCCAG	TTGAAAAAAT	TGAAGAGGTG	ACTGAGGAAG	1500
GTCCACCACT	TTACAAAGTC	CCTGCTAAGG	CCCTAATTTG	AT		1542

(2) INFORMATION FOR SEQ ID NO:67:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1321 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:67:

ATCGAATTAC	TTCAACTCCA	ACTTTACTCT	CAATAAAAAT	CAAATGTAAA	AAGAGGAGCT	60
AAATTTATCT	TTTTCTCCTC	CTTCATCGTT	CTTACTTTTG	ACCATAATAA	GCATTTGGTC	120
CATGTTTACG	TTGGTAGTGT	TTTTCTAGTA	TGTACTGGGG	AGCAGGTTCA	ACTCTTGGAT	180
TGATTTGTTT	TGTAAAGCGA	TTCATCTTTG	ATACTTCCTC	TAGTACGACA	GAGTGATAAA	240
CAGCATCTCT	TGGATTTTTG	CCCCAGGTGA	ATGGACCGTG	ATTGCGTACA	ACAATTCCTG	300
GTACTTCAAC	CGGGTTAAGT	CCGCGATGTT	CAAACCTCTC	TACGATAACC	AGGCCAGTAT	360
CTTTTTTCATA	GGCCACTTCT	ACTTCGTCCCT	TGGTCAAACCT	ACGGGCGCAA	GGGATTGAAC	420
CGTAGAAATA	ATCTGCATGG	GTTGTTCCTG	AGAAAGGAAT	ATCACGACCT	GCCTGAGCCC	480
AAGCAACAGC	TTCTGTGCGA	TGGGTGTGAA	CCACACTACC	AATTTCTGAC	CAAGCCTTAT	540
ATAATTGCAC	ATGAGTTGGG	AAGTCGGAAG	ATGGTCTTAA	ATCCCCCTTAT	AGGATCTTAC	600
CATCTAGATC	AGTCACTACC	ATGTTTTCAG	GTGTCAATTC	GTCATAATCC	ACGCCTGATG	660
GTTTGATAAC	AATGACACCG	AGTTCGCGAT	TGACTTCAGA	TACATTCCCC	CAGGTAAATT	720
TGACAAGTCC	ATGTTTTTGGC	AATGATTGAT	TGGCATCACA	GACTCGTTTA	CGCATAGCAT	780
TGATTACTTG	ATTCATCTTA	CATCAAACCT	GCTTTCCTTAA	TGAGTGGATA	GAGAAAAGCT	840
TGCGCCTCTT	GAATGGCTGC	GCGTGTTTCT	TCTACTGTTT	CACAATTTTC	AGACCACATT	900
TCGATTAGGA	AAGGTCCATT	ATAATTGGTT	TCCTTTAAAA	TATCGAAAGC	TTCTTCCCAT	960
TTGACACAAC	CTTGCCCAA	AGGTACATCT	CGGAACCTGGC	CCTTTGAACT	TTCTGTCACT	1020
GCATAAGTAT	CCTTGAGATG	GAGAGTTGCG	ATGGCATGAT	GACCAAGATA	AAACTCACTA	1080

TAGATATCAT	TATGCCATGC	AGACACATTA	CCAATATCTG	GATATACAAA	GAGGAAGGGA	1140
GAGTCAATCT	CTTTTCTAT	AGCCAAATAT	TTTTCGATGC	TATTGATGAA	AGGATCATCC	1200
ATAATTTCAA	TAGCAAGTAC	CACCTGAGCT	TCTTCAGCCC	AGTCACAGGC	TTTTCTCAAA	1260
TTTTTGATAA	AACGTTGGCG	TGTCTGGGGT	GACTTTTCCT	CATAGTAAAC	ATCGTAACCA	1320
G						1321

(2) INFORMATION FOR SEQ ID NO:68:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1265 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:68:

TTTTTCTGTT	TTTCGGAGCA	AACTGGGCTC	CAGCCGGTTT	TGGCCTTCTT	TCCTTAGCTA	60
CAGCTGGTTT	AGCTGGCTCA	GATTTTTCGG	CTTTCTTTTC	TGCACTTACT	TTTGGTGCCTG	120
CAGGTTTTGC	TTCTACTTTC	GGAGCAGCTG	CAGGCTTAAA	GCTGGCAGCA	ATTTTTGCAG	180
CGACAGCTTC	TTCCACACTT	GATGAGTGGC	TTTTACATC	CAAGCCCAAC	TCTTTTGCAC	240
GCGCTACAAC	TTCTTTACTT	TCTTTTCCAA	GTTCTTTTGC	GATTTTCGTAC	AATCTTTTCT	300
TAGACAAATC	ATGTCCTCCT	CTTCTATTCC	ATAAGAGACC	TCATTTTCTT	TGTAAATCCA	360
GCATCTGTTA	CAGCCAAAAC	CTTTCTCGAT	TTCCCGACTG	CTATGATTAA	TTCCAGTGTT	420
GAAAACACGG	TTACAATTTT	TACTTGATAA	TAATGACTTT	TATCTTGAAT	CTTCTTGGTC	480
AGATTGGGTC	CAGCATCATG	AGCTAGAAAG	ACCAACTTGG	CCTTGCCGTC	TTGAATGGCC	540
TTGACCACCA	ATTCTTCACC	CGATATGATG	CGCCCTGCTC	GCTGAGCAAG	CCCCAAGAGA	600
TTACTTATCT	TTTGCTTATT	CAAGTCCCAA	CTCTCTTCTT	TTCACTTTGT	GATCCACATA	660
AGCGATCAAC	TCGTCATAAA	AGCTTTCTTC	CACCTTCCATG	CTAAAGCTGC	GGTTAAAGAC	720
CTTCTTCTTT	TTGCCTCTA	GGGCTTCTGC	ATTGTCTAGT	TTGATATAAG	CGCCGCGGCC	780
ATTGGCCTTG	CCCGTAGGAT	CAATAAAGAC	TTGTCCTTCC	TTGTTCTTGA	CAATGCGGAG	840
CAAATCACGC	TTATCAATCA	CTTCGTTAGA	CACAACAGAC	TTGCGCAAAG	GGATTTTTCT	900
TGTTTTTCATC	TTTCCCTCCT	CTAGCAGCTT	TTATTCTTCT	ACAGTATCGT	TTTCTACTTC	960
CAACTCTACT	GAAGCAGCGT	CTTCCATGGC	TTCAAATTCG	CTAGCAGACT	TGATATCGAT	1020
ACGGTAACCA	GTCAAGTGAG	CCGCCAAGCG	CACGTTTTGT	CCACGACGAC	CAATGGCAAG	1080
AGAAAGCTTG	TTATCTGGAA	CAACCACCAA	GGCACGTTTG	CTGTCGTTTT	CATCAAAGAT	1140
AACTTGGTCA	ACCTCAGCAG	GAGCGATGGC	ATTGTAGATA	AATTCAGCTG	GATCTGCTAC	1200
CCACTCGATA	ACATCGATAT	TTTCTTCGAT	TGGTACCATG	CGGTCATTTT	TAGCATCGTA	1260
ACGAG						1265

(2) INFORMATION FOR SEQ ID NO:69:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1305 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:69:

ATAAACCAAA	GGAAGCTGAG	CTCTTTAGTC	CCAGCTTCTT	TTTATATATA	AAATTTTACC	60
CGTGAAAAGA	CAGGGCCTTA	GCAGACTTCT	TTTTTACTTC	G TTCACCCTT	GCTTTTTTCTT	120
TGTATGTTTG	GGCGTTGGCA	GTTGGTTATA	CATAGCTAAA	ATCAGGTCTT	ATAGAAACAT	180
CTTATTATCA	AGTTCCTCCA	CTCAAATCAT	TTCTTTGGCA	CCTTTGTATG	GAAACTCAAA	240
AGAAGATTGG	TCAATCTTAT	CTAAGACTGC	TTGCACGGGT	TTAACTAAAA	GCGATCGTCA	300
TAAATGCCGC	CAATAATCTT	GCCGCGGAAG	TAAAGAATAT	ACTCCCCCAT	CATGGAACGG	360
TAAGTCACAT	CATCTAATCC	TGATAATTGT	TCCAAAACAA	ATTCCAAATA	GTTCTTACTT	420
GATGCCATTT	CTAATCTTCT	AGGCTCTGTT	CAACGATAAC	AACCGTATAG	AGTTCTTGCT	480
TAACCTCGCA	TCCAATTGAT	TAAAGCCCT	GCTTTTCCCA	AAAATGCTGA	GATTGCGGAT	540
TTCCCTTAAC	ATAAGCCAAA	CGTGCCTTTC	GAAAGTTCTT	AGCAAAATAA	GCTAGTGCTT	600
CTGTCACAAT	ATGACTACCA	ATCCCTTTCC	TCTGATAGGC	TTGATCAACC	ATAAACAAAC	660
CAATAAAAAC	AGTCTCCTCA	TCAGGATATG	CATAGACAAA	ATCCATAACA	GCCACAAGGT	720
CAAATCCATT	CCAAAATCCA	ACAAAAAACT	TATCAGCCTT	AGCTTTACCT	TCAGGTAGAC	780
AAAGCATGTC	CTCTTTTACA	GTTGCAAAAT	TTGGCTCTGG	TGGACAATGC	TGAAAATACA	840
GAGGATTACT	TTCATATAAA	GATAAAATAC	TTGGAATATC	CTTTTCAGTT	AGTATCCTAC	900
AACTGTAATA	CTTAGATAGT	TGGTCAATCA	TCTTTTCAAA	TTCGATACTT	TCTTGTGCCC	960
TGTGATTATG	ACACAGGAAG	ATGCACTGAT	CGTCATCAGC	CACATAAAAG	TTCTTTCCAT	1020
CGTGCCTAAT	CGTTGTCTCA	AACCTTTGGA	TAAAACCTTT	AGCCTATACA	ACTGGATTTT	1080
CCTCTCTCAA	AAGTATATTC	TTTTGCAGGC	GAACCTCCTC	AAAATCAGTC	GTGTGCAACT	1140
TCAGTAGAAT	ATTCATAGGC	TCGGATAATC	TGAGCGACAA	CAGGATGGCG	AACCACATCC	1200
TTGGCTGAAA	AATGAACAAA	GTCAATCTGA	TGGATGTTCT	TGAGTTTCTC	TTGAGCATCA	1260
ATCAAACCGG	ACTTGACATT	ACGTGGCAGG	TCAATCTGAC	TAATA		1305

(2) INFORMATION FOR SEQ ID NO:70:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1742 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:70:

CTAATCTCCT	TAAAACGTGA	TCTTTTCAAG	AATATTTT	TCTAAACAAT	CCAGCAAGTC	60
TTGGTAAGAA	TAGACTTCGT	AAGTCGGCTG	GGCTTGTGTG	TGATTTTCGA	GGTGATGAGG	120
ATTATAACCAG	ATAGTGTCAG	TCCCCGCATT	ATTGCCACCT	TGAATGTCGG	CGGTTAGAGA	180
ATCTCCAATC	ATCAGCGTCT	TTTCTTTACT	AAATCCAGCA	ATTTGCTGGC	CAATCTTTTC	240
ATAAAAAAGA	GCATCCGGCT	TTTGAGTTTG	CAACTGTTCT	GAGATAAAGA	CTTGATTGAA	300
ATAAGGTGCT	AGACCAGATT	GAGCCAAACG	TCCTGTCTGA	ATGGCAGTAA	TGCCATTTGT	360
CGCAGCATAC	AAGTTATAAT	CACGCTCAAT	GAGGCTGTCC	AAGAGATCAT	GAGCGCCCCG	420
TAGTGTTTGT	CCCTGCTGGG	CGAGGTAAAA	TTGGTAACGC	TGGGCAAGAA	AACCTACCGTC	480
TTTTTCCTGT	CCAAAATGAG	CAAATAAACG	AGAAAAGCGC	GTGTTAACCA	GCTCTTGTTT	540
ACTGATTTTC	TTCAGCTCCA	AGTCTTTCCA	GAGAGCCTTG	TTCATAGGAA	CGTAATAATC	600
TTTATAAGCC	GGAATATCCG	CAACTCCTTC	TTCTTTTAGA	AGTGGAGTCA	AAGCCACATC	660
CTCAGCAGCA	TCAAAATCAA	GAAGAGTGTG	GTCGAGGTCG	AAGAGTACAA	ATTTGTAGAA	720
CAATTTGAGG	TTTTCCCTTC	TGAAAATTCA	TTAAGAACAT	TATATCATAA	AGCACCTCAT	780
ACAATTAAC	AATTTAATCA	CTTAAAAAAA	ATTCGAACAC	TTTCTATACA	ACTGACAGCT	840
CAAATCTTTC	AGAATAGAAC	AATACTAACT	ATCGAACACC	CCGTCTTCAT	AAATACATAT	900
GTAATTCTAG	GCCTAGAATT	CCTATAAACT	AAATGCTTTC	ATACTCTTCC	AAGTAATTGA	960
TTGCCTTAAA	TTTTAATTTT	TGAAGGTTTC	TAAAGCTAGA	ATAGCCCCAT	CACAATCAGT	1020
TTTGATTGAT	TCACAATTTA	GAAACACTAT	AGTTTCACTC	CTGTTAAAAA	AAAAGGAAC	1080
TGCATAAAGC	AATCCCTTTC	TGATTTTGAA	ATCATTTACT	TAACATTTTA	TAGTTGAGAT	1140
AATCAATAGC	TTATCTATAA	AAAGAGTTAT	AGTAAATTC	CTTATTTATT	GATTCCAAGC	1200
TCCGCTAACT	GTATTTGAAT	AACTGACAGT	TCTGCACCAG	CCTGAAAAAG	AGCAGCTGCA	1260
TTATAGGCAC	CTTCTACAAT	TGGAACCCTG	TTGATGATGA	TACTTTTATC	ACTGAAATCA	1320
GTCACCATTT	TTAAGTTCAT	TTTAGCAGAA	CCTAGGTCAA	AAAAGGCAAG	TAAAGTATCT	1380
GCTGGATTTT	CGGAAACAAC	CCTATCTACT	TGATCAAAAC	TCGTTCCAAT	TCCTCCGCCC	1440
TCGGTTCCCTC	CTACATAAGT	AATCGGAACA	TCTTTAGCTA	CTTTACTAAT	CAGTTCAACA	1500
ACACCTTCTG	CAATGTGTTT	GGAATGTGAA	ACGATAACAA	GACCAATACC	AATACTTTCC	1560
ATCAAACCAC	TCCAGTTTCT	AAAATAGCAG	TAAAGAGTAA	TCCTGATGAG	AATGATCCAG	1620
GATCAATATG	TCCAAGAAAC	CACATGCTCC	TAAGACAAGA	GCTAACAGAC	TGGCCATCAA	1680
TAATAGTATT	GTTCTTTTTT	TCATCATTAC	TCCTTAACTA	GTGTTTAACT	GATTAATTCTG	1740
AT						1742

(2) INFORMATION FOR SEQ ID NO:71:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1136 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:71:

GTGGAATGCG	GGGACGCCTT	GTCTAATTTT	GGATCAAGCC	CTGAGTTTGA	CACAGGGAAA	60
------------	------------	------------	------------	------------	------------	----

TGAGCTGGAC	GGACTGCTAT	CTCTGAAGAA	ATTACTGGCA	CCATTAGCCT	ATCAGCCTTG	120
GATGATTATG	TGGCGGCCTT	GTCTCAACAG	GATGTTCCCA	AAGCTTTGTC	TTGCTTGAAT	180
CTTCTTTTTG	ACAATGGTAA	GAGCATGACT	CGTTTTGTGA	CCGATCTTTT	GCACTATTTA	240
AGAGACTTGT	TAATTGTTCA	AACAGGGGGA	GAAAATACTC	ATCATAGTTC	AGTCTTTGTA	300
GAAAATTTGG	CACTTCCTCA	AAAAAATCTG	TTTGAAATGA	TTCGCTTAGC	AACAGTGAAT	360
TTAGCAGATA	TTAAGTCTAG	TTTGCAGCCC	AAGATTTATG	CTGAAATGAT	GACCGTCCGT	420
TTGGCGGAAA	TCAAGCCCGA	ACCAGCTCTA	TCAGGAGCGG	TTGAAAATCG	AATTGCTACG	480
CTGAGACAGG	AAGTTGCCCG	TCTCAAACAA	GAGCTTTCTA	ATGCAGGTGC	GGTTCCTAAA	540
CAAGTTGCAC	CAGCTCCTAG	TCGACCAGCT	ACGGGCAAAA	CAGTCTATCG	TGTCGATCGC	600
AATAAAGTGC	AATCTATCTT	ACAAGAGGCC	GTCGAAAAATC	CTGATTTAGC	ACGTCAAAAT	660
CTAATTCGTT	TGCAGAAATG	CTGGGGAGAG	GTAATTGAAA	GTCTAGGTGG	GCCGGACAAG	720
GCTCTGCTAG	TTGGTTCTCA	ACCGGTTGCT	GCCAATGAAC	ACCATGCTAT	TCTTGCTTTT	780
GAGTCTAACT	TCAATGCTGG	TCAAACATATG	AAACGAGACA	ATCTCAATAC	CATGTTTGGT	840
AATATCCTCA	GTCAGGCGGC	AGGTTTTTCA	CCTGAGATTT	TAGCTATTTT	CATGGAGGAA	900
TGGAAAGAAG	TTGCGCAGC	CTTTTCAGCC	AAAGCCAAAT	CTTCTCAAAC	TGAAAAAGAA	960
GTAGAAGAAA	GCCTGATTCC	AGAAGGATTT	GAATTTTTTG	CTGATAAAGT	GAAGGTAGAG	1020
GAAGACTAAA	GAAAGATTTT	ATGATACAAT	AAGTTTATGA	ATAACAACA	ATTTATTATT	1080
ATGGCGCTAT	TTACAGCTGC	TGAGACCTAT	TTTTTCAATG	AAGCCTGGAT	GAATGG	1136

(2) INFORMATION FOR SEQ ID NO:72:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1670 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:72:

CTGTCTCTGA	AACAGTCACA	TCAAGTGCCT	CTGAACAANC	GCCCCNCCTA	GGTNGACGGT	60
ATCGATAAGC	TCGATCTGTG	ATTTTCAGAGA	AGAAATCAAG	TGCTGTAACA	GAAGTAAGAT	120
GTAATTGTAT	GTAAAGGAGA	CGTCATGTTA	AATAGTATTG	TAACCATTAT	TTGTATTGCC	180
CTTATCGCGT	TTATCTTGTT	TTGGTTTTTC	AAAAAGCCTG	AAAAATCTGG	ACAAAAAGCC	240
CAGCAAAAAA	ACGGATACCA	AGAGATTCGA	GTGGAAGTCA	TGGGAGGCTA	TACTCCTGAG	300
TTGATTGTCC	TCAAGAAATC	AGTGCCAGCC	CGCATTGTCT	TTGACCGCAA	GGATCCTTCA	360
CCATGTCTGG	ATCAAATTGT	TTTTCCAGAT	TTTGGTGTAC	ATGCGAACCT	GCCAATGGGG	420
GAAGAGTATG	TAGTGGAAT	CACGCCTGAA	CAGGCTGGAG	AGTTTGGCTT	TGCTTGTGGT	480
ATGAACATGA	TGCACGGCAA	GATGATTGTA	GAGTAGGTGG	AGACTATGAC	AGAAATTGTG	540
AAAGCAAGCT	TAGAAAAATG	CATTCAAAAA	ATCCGTATCC	GAGCTGAAAA	AGGCTATCAT	600
CCAGCCCATA	TCCAGCTTCA	AAAGGGAATT	CCAGCTGAGA	TTACCTTTCA	TTCGTGCTAC	660
TCCTTCAAAC	TGTTATAAGG	GAAATTCTGT	TTGAAGAAGA	AGGTATCTTG	GAAGCAATCG	720
GCGTAGATGA	GGAGAAAGTC	ATTCGTTTTA	CACCTCAAGA	ATTAGGGAGA	CATGAATTTT	780

CTTGTGGCAT	GAAGATGCAA	AAGGGAAGCT	ATATAGTCGT	TGAGAAGACT	CGAAAATCTC	840
TATCTCTCCT	GCAAACGTTT	TTGGATTACT	AGTATCTTTA	CTGTGCCTCT	TGTGATTCTC	900
ATGATTGGGA	TGTTGGCAGG	TAGCATTAGT	CATCAAGTCA	TGCATTGGGG	AACCTTTTTTA	960
GCAACAACGC	CTATTATGTT	AGTTGCGGGT	AAGCCATATA	TCCAGAGTGC	TTGGGCCAGT	1020
TTTAAAAAGC	ACAATGCCAA	CATGGATACC	TTGGTTGCGC	TGGGAACTCT	AGTGGCTTAT	1080
TTCTATAGCC	TAGTTGCTCT	CTTTGCTGGT	CTCCCTGTTT	ACTTCGAAAG	TGCTGGATTT	1140
ATCCTCTTTT	TCGTTCTTTT	GGGAGCAGTT	TTTGAGGAAA	AAATGAGGAA	AAATACGTCC	1200
CAAGCTGTGG	AGAAATTACT	GGACTTGCAA	GCTAAAACCG	CAGAAGTCTT	GAGTGATGAT	1260
AGTTATGTCC	AAGTTCCTTT	GGAACAAGTC	AAGGTACGCG	ACCTTGATTG	CAGTGCCTCC	1320
CGGTGAAAAG	ATTGCTGTTG	ATGGTGTCGT	AGTAGAAGGT	GTCTCTAGTA	TTGACGAATC	1380
CATGGTGACA	GGTGAGAGTC	TGCCTGTGGA	CAAGACAGTT	GGAGATACTG	TCATTGGCTC	1440
AACCATCAAT	CATAGTGGA	CGCTTGCTCT	TAGAGCAGAA	AAAGTTGGCT	CAGAGACTGT	1500
TTTGGCTCAG	ATTGTAGATT	TTGTGAAGAA	AGCTCAGACA	AGTCGTGCGC	CGATTCAGGA	1560
CTTGACGGAT	AAGATTTTCT	GGATTTTGTG	CCCAGTAGTT	GTCATTTTAG	GAATCATGAC	1620
CTTTTGGGTT	TGGTTCGTCT	TGCTCAGGGA	TAGTGTGGTC	GTGCTTGGAG		1670

(2) INFORMATION FOR SEQ ID NO:73:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1252 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:73:

ACAAGAACAA	TTGGAACAGG	TACAGGCTGT	TAAAAAATCG	ATTAACACAG	CTAGTGAAGA	60
AGTGAAAAAC	CAAGTCTTGC	TACCCATGGC	TGATCACTTA	GTGGCTGCTA	CTGAGGAAAT	120
TTTAGCGGCT	AATGCCCTCG	ATATGGCAGC	GGCTAAGGGG	AAAATCTCAG	ATGTGATGTT	180
GGATCGTCTT	TATTTGGATG	CAGATCGTAT	AGAAGCGATG	GCAAGAGGAA	TTCGTGAAGT	240
GGTTGCCTTA	CCAGATCCAA	TCGGTGAAAGT	TTTAGAAACA	AGTCAGCTTG	AAAATGGTTT	300
GGTTATCACA	AAAAAACGTG	TAGCTATGGG	GGTCATCGGT	ATTATCTATG	AAAGCCGTCC	360
AAATGTGACG	TCTGATGCGG	CTGCTTTGAC	TCTTAAGAGT	GGAAATGCGG	TTGTTCTTCG	420
TAGTGGTAAG	GATGCCTATC	AAACAACCCA	TGCCATTGTC	ACAGCCTTGA	AGAAGGGCTT	480
GGAGACGACT	ACTATTCATC	CAAATGTGAT	TCAACTGGTG	GAGGATACTA	GCCGTGAAAG	540
TAGTTATGCT	ATGATGAAGG	CCAAGGGCTA	TCTAGACCTT	CTCATTCCTC	GTGGAGGAGC	600
TGGCTTGATT	AATGCAGTAG	TTGAGAATGC	CATTGTGCCT	GTTATCGAGA	CAGGAACTGG	660
GATTGTCCAT	GTTTATGTCT	ATAAGGACGC	AGATGACGAC	AAGGCACTGT	CTATCATCAA	720
CAATGCCAAA	ACCAGTCGTC	CTTCTGTCTG	CAATGCCATG	GAGGTTCTGC	TGGTTCATGA	780
AGACAAGGCA	GCAAGCTTCC	TTCTTCGCTT	GGAGCAAGTG	CTGGTTGCAG	ATCGAAAAGA	840
AGCTGGGTTG	GAACCAATTC	AATTCCGCCT	AGATAGCAAA	GCAAGCCAGT	TTGTTTCAGG	900
TCAAGCTGCT	CAAGCACAAG	ACTTTGATAC	CGAGTTTTTA	GACTATATTC	TAGCTGTTAA	960

GGTTGTGAGC	AGTTTAGAAG	AAGCGGTTGC	GCATATTGAA	TCCACAGTAC	CCATCATTCG	1020
GATGCTATTG	TGACGGAAAA	TGCTGAAGCT	GCAGCATACT	TTACAGATCA	AGTGGACTCT	1080
GCAGCGGTGT	ATGTTAATGC	CTCAACTCGT	TTCACAGATG	GAGGACAATT	TGGTCTTGGT	1140
TGTGAAATGG	GGATTTCTAC	TCAGAAATTG	CACGCGCGTG	GTCCAATGGG	CTTGAAAGAG	1200
TTGACCAGCT	ACAAGTATGT	GGTTGCTGGT	GATGGGCAGA	TAAGGGAGTA	AG	1252

(2) INFORMATION FOR SEQ ID NO:74:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1785 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:74:

CTGCCCTAGC	AGGAACGCAA	GAAGGAACTG	GAGAATAGGC	ATTTTCAAAA	TTATAACCTA	60
CACTAGCCAT	CATATCTAAT	GTTGGAGTGC	TAAGTAGCTT	ATCCTTACTA	TTCAAGGATA	120
AGGCGTCTGC	TCTCATTTGA	TCTACAACAA	TCAAAATAAT	ATTTGGTTGT	TTTGTCTGAA	180
CCATAAAATC	TCCTTTCTAA	TATGGCAAAA	GAGGCACAAG	AAGATATCTA	CCTTTACTGC	240
ACCCCTTTCT	ATATCAATCT	CTCTATATAA	AGCAATAACA	TTCTTGTTAT	GTTTTATAGA	300
ACAATGGACT	AAAATATGAC	TAAATCGATT	AGGAAATTCA	AATCATTTTC	TAGTACTGTT	360
TTAGTAAGTT	ACAGTGTAAT	ATTCCAACCT	CAATAAATTA	TAAACCTTTG	TCTAATAACA	420
ATTTTAGTGG	AGATAAGAAA	TCCTACACCT	AACTCATCTT	ACACGTAATC	TATTTCTATT	480
TTATCACAAA	AAACGCAAGT	AAGACCATTA	ACTCAATTCA	GTTTTATCTG	CCATTTTCAC	540
AAATGGGAAA	TAAGTCAAGA	CACTAATAAT	CAAACAAACA	ACTGATAAGA	TGATGGCACG	600
CCAATCAAAT	GCTGTAGAGA	AGAAACCATA	TAAAAATTGGA	GGCATTACCC	AAGTAACATT	660
TTGTGTAACA	GGTGAAACAA	GACCCCAGCT	TGTTGCCCCAG	TAAGCTACCG	TTGCCATGAA	720
AACCGGGCTA	AGTACAAATG	GTATAAATAG	CAAAGGATTC	AAGACAACCTG	GTAAACCATA	780
ATTCGATACC	GGCTCACCAA	TATTAAACAG	AACTGGTGCT	AGACCAAGTT	TAGCAACTTT	840
TCGATAATGA	CTGTTTCTTG	AAAAAATTAA	AATAGCAAGT	ACTAATCCTA	ATCCTCCAAA	900
CCAGACAAAC	GCCCCAAAAG	ACCCACTTGT	CCATATATAA	GGAATCGGTT	CACCTTTTTG	960
GAAAGCATCC	AGATTCGCTA	ACATAGCAAC	TCCAAATAGC	CCTTCCATGA	TGGGAGCCAA	1020
TACATTTCTT	CCATGGAGAC	CAAAAAACCA	GAATAACTTA	TTCAAAAAGA	TCATCAGAAT	1080
AACTGCAAAG	AACTTTTGAG	ACAAACCTAG	TAATGGCGTT	TGTAACACCT	TGTAAACCCA	1140
ATCAATCAAT	AAGTCATTGC	TAAGTAAATG	GAAAACATAA	GTCAAGATGG	CTACTATATA	1200
CATCGCCATA	AATCCTGGAA	TGATAGAAGT	GAACGGCTTA	GCAATCGCAG	GGGGAACCTGA	1260
ATCTGGTAAC	TTGATTACCC	AGTTCTTTTT	CATTACTTTA	CAGAAAATAA	TAGAGGCTAA	1320
AAATCCAATC	ATCATGGCTG	TAAAGTAGCC	TCTGGCATT	ATATGGTTTC	CTGGAATCAC	1380
ATTCCCAATA	GTTACCATCA	GATTTTTTACC	ATCAAATGCT	AGATTATCAA	TTCCATGTTA	1440
AGATTTGATC	TAATTTTACA	TCTCCTACAT	TTGCCAAAGG	GAAACTCTTT	GTAAGTGTAC	1500
TTCCAATCGA	AATGACAAAC	GAAGCAAGTG	ATACCAAACC	AGCAGAAACT	GTATCAACCT	1560

TGTAAATCTT	AGCGATATTC	ACTCCCAAGC	AATAGATGAA	CAACAAGGAA	ACAATTGGTA	1620
TACTTCCCTT	GAATACCAAA	TTATTGATGT	CAACAAGCCA	CTGAAAGGTT	TTCGTAATAC	1680
TTCTTAGGTG	AAATTGTTGT	GGTAAATCCA	CTAGAAAAGC	ATTTAATAAC	AAAGCAATGG	1740
AACCTGTCAT	AATAACAGGC	ATAGTCCCCA	CAAATGAATC	ACGTT		1785

(2) INFORMATION FOR SEQ ID NO:75:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1386 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:75:

ATCGAATTTT	ATTTCTATTT	CCTATTCCAT	TTTTATTCAA	AAAATCAAAA	AGCAAAC TAG	60
AAAGCTGGTC	GCTGGTGGTT	CAAAACACTG	TTTGAGATT	GTCAATAGAA	CTGACAAACC	120
CTGTAATATA	CCTGCATATA	TACATACGAC	AAGGCGATAC	TACCCTAGTT	TGAAGAGATT	180
TTCGAAGAGT	ATTCATTTTT	GTCTTTTACT	TATTATACCA	TATTCACATA	AAAAAACGAA	240
CATTCTTATC	CTAAAAAATG	CTCATTTTTT	TTAAATTATC	AATCTAAATC	TGGTTTATAG	300
AAGGAACGAT	TATCCATAGC	GAAGATTTTA	TTGGTCATCT	CTCCTTTATC	CACCAAAGCC	360
AGAGCTGTTG	ACATCATCAT	CATGCTTGCA	TCCAGATTGT	CAATCATATG	GATAATCTCT	420
GCCTCCATAA	TACGTGGACG	GACTGGAATT	TCCATATTCA	AGCAAGCCGT	GGTGGACTTG	480
AGGATGACAT	GACGAAGCAA	AACGACTTCT	TCCTTGGTAT	CATCGATGCC	GAGTTCCATA	540
ACTGTCTTGG	TAATTTTCGCT	ATCAATGAGA	GCGATATGTC	CAAGAAGATT	ACCTCGCACT	600
GTGTACTCTG	TCTGGTCTGG	CCCCGTCAAC	TCGATAACCT	TAGCTAAGTC	ATGCAGCATA	660
ATCCCCGCAT	AGAGCAGGCT	CTTATTGAGC	TGAGGATAAA	CTTCGCTAAT	AGCGTCTGCC	720
AAACGTACCA	TGGTCGCCGT	ATGATAAGCC	AACCCCGTTT	CAAAGGCATG	GTGGTTGGTC	780
TTGGCGGCTG	GATAGGAGTA	GAATTCCTTA	TCATACTTGG	TGTAGAGATT	TCGGACAATC	840
CGTTGCCAGA	CAGGATTTTC	AATTTTGAAA	ATCATTTGCG	ACATGTAGTC	ACGAATTTCC	900
TTGACATCAA	CTGGTGACTT	GACCTTGAAA	TCAGCTGGGT	CATTGGGTTT	ACCAGCTTGA	960
GGCAGGCGGA	GAGTAATTTG	ATTGACTTGA	GGGGTATTGT	TATAAACTTC	TCGGCGTCCT	1020
TTCATGTGGA	CAACCTTACC	TGCGGTAAAG	GCCTCAATGT	TATGAGGTTG	GGCATCCCAG	1080
AGCTTCCCAT	CAATCTCGCC	ACTATCATCT	TGGAAGGTAA	AGGCTAGGTA	GTTTTTCCCA	1140
GCTCGAGTTT	GCCTCAGGTC	AGCTGATTTG	ATTAGGTAAA	AGCCTTCAAA	TAATCATCT	1200
TTTTTCATGT	GACTAATCTT	CATATTCTTC	CTCATTTTCT	TGAAAATGGA	GTAGATCAAG	1260
CGCAGGCTCA	CCTTCTGACA	ACTCAATGTG	ACGGAGCGTC	CGCTCGATAG	CTATGGTACG	1320
ACGGTTTAAT	AATTCGATCA	ATATTGCCAG	AGGCATGTTG	GAGATGTTTT	TGTGCCTTGA	1380
CCAGAA						1386

(2) INFORMATION FOR SEQ ID NO:76:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1167 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:76:

```

CTCAGATTAC AGAGGACAAT CAACTGGTTC ATTTTCGTTT CCAGTTTCAA AAAGGCTTAG      60
AAAGGGAGTT CATCTATCGT GTGGAAAAAG AAAAAAGTTA AGGCAGGTGT TCTCCTCTAC      120
GCAGTCACCA TAGCAGCCAT CTTTAGTCTT TTGTTGCAAT TTTATTTGAA CCGACAAGTC      180
GCCCCACTATC AAGACTATGC TTTGAATAAA GAAAAATTGG TTGCTTTTGC TATGGCTAAA      240
CGAACCAAAG ATAAGGTTGA GCAAGAAAAGT GGGGAACAGG TTTTAAATCT AGGTCAGGTA      300
AGCTATCAAA ACAAGAAAAC TGGCTTAGTG ACGAGGGTTC GTACGGATAA GAGCCAATAT      360
GAGTTTCTGT TTCCTTCAGT CAAAATCAAA GAAGAGAAAA GAGATAAAAA GGAAGAGGTA      420
GCGACCGATT CAAGCGAAAA AGTGGAGAAG AAAAAATCAG AAGAGAAGCC TGAAAAGAAA      480
GAGAATTCCT AGTCAATTCA ACTATAATGC GTTGAATCCA GAATAGTCCA CTGTAGTTTC      540
TAGAAAATTG CTGGAAATGG ATGTTAAGCT CCAATTCATT TGTTTATATC TTATTTTCAGT      600
CCACTATACT TTGTGCTAAA TTAAAGATAT GAAACATGAT TTTAACCACA AAGCAGAAAC      660
TTTCGATTTT CCTAAAAATA TCTTCCTCGC AAAC TTGGTA TGTCAAGCAG CCGAGAAACA      720
GATTGATCTT CTATCAGACA AAGAAATTTT AGATTTTCGGT GGTGGCACGG GTCTATTAGC      780
CTTGCCCCTA ACCCCTAGCC AAGCAGGCTA AGTCAGTCAC TCTTGTAGAC ATTTCTGAGA      840
AAATGTTGGA GCAAGCTCGT TTGAAAGTGG AGCAGCAAGC AATCAAGAAT ATCCAGTTTT      900
TGGAGCAAGA TTTACCGAAA AATCCCTTGG AGAAAGAGTT TGATTGCCTT GCTGTTAGTC      960
GGGTTCTTCA TCATATGCCT GATTTGGATG CGGCTCTCTC ACTGTTTCAT CAACATTTGA     1020
AGGAAGATGG GAAACTCATC ATTGCTGATT TTACCAAGAC AGAAGCTAAT CATCATGGAT     1080
TTGATTTAGC TGAAGTGGAA AACAAGCTAA TTGAGCATGG GTTTTTTCATC TGTGCATAGT     1140
CAGATNCTCT ATAGCGCTGA AGANCTG                                     1167

```

(2) INFORMATION FOR SEQ ID NO:77:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 916 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:77:

```

TCTCCCAACA TATAATTTCC GTTTTCCAAT CCCCAGCTG TCATACAGTC TGTGATAAGA      60
GCGATGTTTT CTGTTCTTTT TTGTTTGATA AGAATTTTCGC AAGCCTTTGG ATCTACGTGG     120

```

TGACCATCAC	AGATCAACTC	TGCATAGGTA	TGTGGCAATT	GGTACATGGC	TCCAACCATA	180
CCCAATTCAC	GGTGAGTCAA	CCCACGCATT	CCATTGTAGG	CATGCACCCA	AACACTCGCT	240
CCAGCATCGA	CTGCTTTTTT	GGCTTCATCA	AAAGTCGCGT	TTGAATGTCC	AAGAGCAACC	300
GTCACACCTT	CGCCCGTAAC	TGTACGAACA	AAGTCTTCCA	CCCCATCACG	TTCTGGTGCA	360
ATCGAATTTT	ATTAAGCAAG	CCATTTGCCG	CTTTTTTGCCA	AGAATGAAAC	TCCTCAACAC	420
CCGGGTCTCT	CATATAAGTT	GGATTTTGTG	CCCCCTTAAA	AGTTTCTGTG	AAATATGGAC	480
CTTCATAATA	AATCCCACGA	ATCTTAGCAC	CTGTTGCTTC	TTTATAATGG	TTTCCAAGAT	540
TTTCAGTGAC	TGCAAGCAAT	TGCTCATAAG	TGGCTGTAA	AGTTGTGGGT	AAGAAACTGG	600
TAACACCGGT	ACTAAGAAGT	CCTTCACTCA	TAGTATGCAA	TGTACCTTCA	ATGTTGTTGT	660
CCATCACATC	TACACCTGCA	TATCCATGAA	TATGAGTATC	CACAAGACCT	GGGGCAATGC	720
TATAACCTGT	ATAGTCAATC	ACCTCAGCCC	CTTCAGGAAT	CTGCTCTACA	TGTTTCCCAA	780
ACTTGCCGTC	CACAAGTTCC	AAGTAACCAC	CTCGACAAAT	CCGTGTGGGT	AGAAAAACTG	840
ATCCGCTTTA	ATATAGTTAG	GCATAATGTT	AACCTCCTTA	AAAGATTGAT	TCTACAATTT	900
ATTATGTCAA	TTCGAT					916

(2) INFORMATION FOR SEQ ID NO:78:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 786 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:78:

CTGGATTAAA	ACGAGGCAGT	TTCAGACTAA	TATCCAAGTC	GTAAGAAATG	CCTGAAATAA	60
GCTTTTCTAA	ATTGTCCAAA	GCTTGCGGGA	AAACGCTCTT	GGAATAGTTT	CTCTAAAGAA	120
CTTGCTGATA	TAAAGACATC	TTGTCTCGAA	CGCAAGGGAA	CTTCTCTGAG	CGGTAGATTT	180
TCTTTAATCG	CTGTTAAAAC	TTGAAGAACT	TCTCTATCCC	TGCTTTCAAA	AGCGTTGACC	240
CGATAAAGAG	GTAAGATAGG	ATGATGAAAT	TCGCTTGCTA	GTGTTTCTGG	ATAAACCCTT	300
ATATAGTAAT	CACAGCCTAG	TTCTAACGAC	TCAACTCTAT	CAAAATAAGG	CACAATGACC	360
GCGATATCCT	CCAGGTACTG	GGACAGGACT	GACCAAGTTT	TCTCCCCCTG	CATCTTGGCT	420
GTCGAAAGCT	TCATCAACTG	CTGATAGCCC	ACACTAGATA	GAGCTAAAAA	GCGCAAATTC	480
ACTTCCTGAT	CATCTACAAA	CACTGTCAAT	TCAAGCCCTA	GCAAAGGATG	AATGCCGTAT	540
TTTTTTGTAA	TCTCTAGAAA	GTCGAAAGCG	CCATAAAGAT	TGTCAATATC	CATCATAGCC	600
AAATGAGTGT	AGCCGTATTC	TTTAGCTGCT	CTCACATACT	TTTCGATCGA	AATGACGCTT	660
TCCATAAAAC	TATAGACTGT	TTTTGTATCT	AGTTGTGCGA	TCAATTTACA	CTTCTCCTCT	720
ATCCTTCTCA	CTATATTATA	CCATTTTCAC	CTATAAATGG	CTTCTCTTGA	GAAAAATTTT	780
GATCAG						786

(2) INFORMATION FOR SEQ ID NO:79:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1213 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:79:

CACTTTCAGC	TTCTTCTCTT	TTTGAACGGT	TATAAACACG	AATCAGATTC	CCTATTTCTT	60
GCGATTTATG	TGATTCCTTA	TTTTCCAATC	TAAAGTATAG	TGAAATGAAA	TAAAACATGC	120
GCAAATCGAT	TAAGGAATTT	AATCTAATTT	CTAACAATGT	CTTAGAAATC	AAAGTGTACT	180
ATTTTAACTT	CAATGCACTA	AACATCTAAT	ACTCAATAAA	AATCAAAGAG	CAAAC TAGGA	240
AACTAGCCGC	AGGTGGCTCA	AAACACTGTT	TTGAGGTTGT	AGATGAAACT	GACGAAGTCA	300
GTAACCATAC	ATACGGCAAG	GCGACGCTGA	CGTG GTTTGA	AGAGATTTTC	GAAGAGTAGC	360
AAAATGGAAA	AAGGAGTGAG	TGAAGCACAT	CGCCTCCCCA	CTCCTTTTTC	TGTTTTTAGG	420
CTGTTTTTTC	AACCTTCAAG	ATTTTTACAT	CATAGCTACC	AACAGGCGTT	TCAATGGTTG	480
CTGTATCACC	TGTTTTCTTG	CCAATCAAGG	CCTGCCCAAT	TGGGCTTTCA	TTTGAAACCT	540
TACCTGCAAA	GGCATCCGCA	CCAGCTGAAC	CTACGATAAT	ATAAACTTCT	TCTTCGTCCT	600
CACCAATTTT	TTGGATGGTG	ACTGTTTTAC	CAATCGCTAC	TTCGTCCTGG	GCAACTGCGT	660
CGCTATTGAC	GATTTTCAGC	TAGCGGATTT	TTGTTTCTAA	GCTAGAGATT	TGTCCTTCGA	720
CAAAGGCTTG	TTCATCCTTA	GCTGCTTCGT	ACTCACTGTT	TTCTGAAAGG	TCACCGTATG	780
AACGGGCAAT	CTTAATGCGT	TCTACCACTT	CTGGTCGACG	AAACCAATTT	CAATTCCTCT	840
AATTCTTTTT	CAAGTTTTTC	CTTTTCCTCA	AGGGTCATAG	GATATGTTTT	TTCTGCCATT	900
TTTCTCAACT	TTCTTCTGAT	AATATTTTCT	AAAGAAAATT	ATGTGAAAGTA	TCACATAATT	960
TTAGTTTGTT	TAGTTTAATT	TGCTGTTGAC	ATGTTTCAGCG	ACATTGCGGT	CGTGGTCTTC	1020
TTGATTGTTA	GCATAGTAAA	CCTTGCCTTC	TGTGACATCT	GCTACAAAAGT	AAAAGTTATC	1080
GCTCTTAGTT	TGATTGATGC	TTGACTCAAT	CCGCATCCAA	GACTTGGA CT	ATCGACTGGA	1140
CCAGGCATGA	GACCTACATT	TTATAAACA	TTATAAGGTG	AATCAATGTT	GGTATCAATC	1200
GCAACATCCT	CAG					1213

(2) INFORMATION FOR SEQ ID NO:80:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1173 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:80:

TGCGGCTGAG	TTGGGAATTC	CTATCGTTAA	TAAGCGTGTA	TCGGTGACAC	CTATTTCTCT	60
GATTGGGGCA	GCGACAGATG	CGACGGACTA	CTGGTTCCTGG	CAAAAGCGCT	TGATAAGGCT	120
GCGAAAGAGA	TTGGTGTGGA	CTTTATTGGT	GGTCTTTCTG	CCTTAGAACA	AAAAGGTTAT	180
CAAAAGGGAG	ATGAGATTCT	CATCAATTCC	ATTCCTCGCG	CTTTGACTGA	GACGGATAAG	240
GTCTGCTCGT	CAGTCAATAT	CGGCTCAACC	AAGTCTGGTA	TTAATATGAC	GGCTGTGGCA	300
GATATGGGAC	GAATTTATCA	AGGAAACGGC	AAATCTTTCA	GATATGGGAG	CGGCCAAGTT	360
GGTTGTATTC	GCTAATGCTG	TTGAGGACAA	TCCATTTATG	GCGGGTGCCT	TTCATGGTGT	420
TGGGGAAGCA	GATGTTATCA	TCAATGTCGG	AGTTTCTGGT	CCTGGTGTGG	TGAAACGTGC	480
TTTGGAAGAAA	GTTCGTGGAC	AGAGCTTTGA	TGTTAGTAAC	CCGAAAACCA	GTAAAGAAAA	540
CTGCCTTTTA	AAATCACTCC	GTATCCGGTC	CAATTGGTTT	GGTCAAATGC	CCAGTGAGAG	600
ACTGGGTGTG	GAGTTTGGTA	TTGTGGACTT	GAGTTTGGCA	CCAACCCCTG	CGGTTGGAGA	660
CTCTGTGGCA	CGTGTCCTTG	AGGAAATGGG	GCTAGAAAACA	GTTGGCACGC	ATGGAACGAC	720
AGCTGCCTTG	GCCCTCTTGA	ACGACCAAGT	TAAAAAGGGT	GGAGTGATGG	CCTGTAACCA	780
GGTCGGTGGT	CTATCTGGTG	CCTTTATCCC	TGTTTCTGAG	GATGAAGGAA	TGATTGCTGC	840
AGTGCAAAAT	GGCTCTCTTA	ATTTAGAAAA	ACTAGAAGCT	ATGACGGCTA	TCTGTTCTTG	900
TTGGATTGGA	TATGATTGCC	ATCCCAGAAG	ATACGCCTGC	TGAAACTATT	GCGGCTATGA	960
TTGCGGATGA	AGCAGCAATC	GGTGTTATCA	ACATGAAAAC	AACAGCTGTT	CGTATCATTC	1020
CCAAAGGAAG	AGAAGGCGAT	ATGATTGAGT	TTGGTGGTCT	ATTAGGAACT	GCACCCGTTA	1080
TGAAGGTAA	TGGGGCTTCG	TCTGTGCACT	TCATCTCTCG	CGGTGGACAA	ATCCCAGCAC	1140
CAATTCATAG	TTTTAAAAAT	TAAGAAAATA	GGA			1173

(2) INFORMATION FOR SEQ ID NO:81:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1209 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:81:

TCGGAATCTG	AGCTAGTGTA	GCTTCCTTAA	TCTTATCTGA	TAAGATAGCT	GTCATATCAG	60
ACTCAATCAT	TTCCTGGAGC	AATCAACATT	GACTCGTATA	TTCCGACTAG	CGACCTCGCG	120
TGCCACAGAC	TTGGTAAAGC	CAATCAAGCC	AGCCTTAGAA	GCAGCATAGT	TAGCTTGACC	180
AATATTCCCC	ATCAAACCAA	CAACACTAGA	CATATTAATG	ATAGCACCTT	CTCTGGCCTT	240
CATCATCGGT	TTCAAGACTG	ATTGTGTCAT	ATTAAAGGCA	CCAGTCAGAT	TGACCTTGAG	300
CACTTTTTCA	AAATCTGCTT	CTGTCATCTT	GAGCATAAGA	GTATCTTGGG	TAATCCCTGC	360
ATTGTTGACC	AAAACATCTA	CTGAACCCAG	TTCTGCAATA	GCTTGATCAA	TCATACGCTT	420
AGCGTCTGCA	AAATCTGATA	CATCTCCTGA	AATGGGAACC	ACCTTGATAC	CATAGTTTGA	480
AAACTCAGCG	AGCAATTCTT	CTGAGATTGC	CCCACGACTG	TTTAAGACAA	TGTTGGCTCC	540
TGCTTGAGCA	AACTTGTTGGG	CGATGGCAAG	ACCAATTCCA	CGACTCGAAC	CTGTAATAAA	600
GATATTTTTA	TGTTCTAGTT	TCATTTTTTT	CCTTTCAAAA	CTTCTACTTA	TTTTAGTCTA	660

TTTTTCTAAA	AGTGCTACTA	AACTCGCTTG	ATCTTCCACA	TGAGCTAAGT	GAGCAGTTTG	720
ATCAATTTTT	TTAACAAAAC	CTGACAAGAC	TTTCCCCGGT	CCAATCTCGA	ATAAAGTTGC	780
TTATGCCTGC	TTCTTGCAATG	ACCCCAATAC	TTTCATAGAA	ACGAACGGGT	TCCTTGACCCT	840
GACGCGTCAA	GAGCTGAGCA	ATGTCCTCTT	TTTGCAATCAC	AGCAGCTTCT	GTATTGCCGA	900
CTAGGGGACA	AGTAAAATCT	GAAAAACTTA	CCTGAGCTAG	AGTTTCAGCT	AGTTTCTGGC	960
TAGCAGGCTC	AAGGAGAGCG	GTGTGAAAGG	GACCTGACAC	CTTAAGAGGA	ATCAAGCGTT	1020
TGGCACCTGC	TTCTTGCAAA	AGTTCAACCG	CTCGATCAAC	TGCAACCACT	TCTCCAGCAA	1080
TGACGATTTG	TGCAGGTGTG	TTATAGTTGG	CTGGAGTAAC	CACTCCAAGT	TCCAGAAGCT	1140
TTTGTACAGG	CTTCTTCAAT	GACCTCTACT	GGCGTATTGA	GAAGTGCTAC	CATCTTGCCA	1200
AGTTCAGCA						1209

(2) INFORMATION FOR SEQ ID NO:82:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 813 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:82:

ATGACACGTC	TGTTCTCTCA	AGCAGAAATG	GCAGAGTAAC	AAGCTCGATA	TTGAGGTAGC	60
CGATAAAGAA	TTGGCTGAAT	TTGAAGCTCA	GATTAAACAG	GAAGTGAAG	CTCCAACCTG	120
TAGTGAGTCC	TCAGGTTGAA	GAAGAGCCTC	AGCTCATCCA	GTTGGCCCAA	TGTATGAAGA	180
ACCAGAAGTA	AATCCAGTGC	ATCCGACAGG	TCCAACACCA	GCTACAGAAA	CTGTTGATTC	240
AATACCGGGA	TTTGAAGCAC	CGCAAGAATC	TGTTACAATT	TTATAAGAAA	TATTCTGAGA	300
ACAATATCTT	ATCCTTATAT	TTCCAGCGAG	CAGGAAATGG	TGTGAGTCCT	GCATTCCCTA	360
TCGATAAGAT	TATCCTCTCA	AACTATCAAG	TCTGAATCTA	GTAAGATTTG	ACGTTCCCCA	420
CGTTACGGGA	TAAGAGAGAG	AAAGACTAAA	TCTTTTTTCCG	AATAAAGGTG	GTACCACGAT	480
TTTCGTCTTT	TTTGGAAGTC	GTGGTTTTTA	ATTTGTTATT	ATTATAAAG	GAGATACCAT	540
GAAACTCAAA	GACACCCTTA	ATCTTGGGAA	AACTGAATTC	CCAATGCGTG	CAGGCCTTCC	600
TACCAAAGAG	CCAGTTTGGC	AAAAGGAATG	GGAAGATGCA	AAACTTTATC	AACGTCGTCA	660
AGAATTGAAC	CAAGGAAAAC	CTCATTTTAC	CTTGCAATGAT	GGCCCTCCAT	ACGCTAACGG	720
AAATATCCAC	GTTGGACATG	CTATGAACAA	GATTTCAAAA	GATATCATTG	TTCGTTCTAA	780
GTCTATGTCA	GGATTTTACG	CGCCATTTAT	TCC			813

(2) INFORMATION FOR SEQ ID NO:83:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 953 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:83:

ATCGAATTAT	TTTGAAACAA	GGTGGATCAG	CTATTTTGGC	CTTGATTAGT	ATTTTACTCT	60
TTAAATACAC	TTGAAGGTCG	ATTCTAATCT	CGCTAATCCT	TTTAAATCCA	GAATAAGGGA	120
AATATGTTAT	ACTTGTTTTT	AAGAAAAAAG	TTTCATTGAA	TTGGTTTTGA	GGAGTTAGAA	180
ATGAAAGTAT	TAGTGACAGG	TTTTGAGCCC	TTTTGAGGCC	ATTAAAGGTT	TACCAGCTGA	240
AATCCATGGT	GCTGAGGTCC	GTTGGCTAGA	GGTGCCGACA	GTTTTTCACA	AATCTGCTCA	300
AGTATTGGAA	GAAGAGATGA	ATCGTTATCA	ACCTGACTTT	GTCTTTTGTA	TTGGGCAAGC	360
TGGTGGAAGA	ACTAGTTTGA	CACCTGAACG	AGTGGCCATT	AATCAAGACG	ATGCACGTAC	420
TTCTGATAAC	GAAGATAATC	AACCGATTGA	CCGTCCCATT	CGCCCAGATG	GTGCTTCGGC	480
CTACTTTAGT	AGTTTGCCGA	TTAAAGCGAT	GGTTCAAGCT	ATAAAAAAGA	AGGATTACCG	540
GCCTCTGTTT	CCAATACGGC	AGGGACTTTT	GTCTGCAGCC	ATTTGATGTA	TCAGGCTCTC	600
TATTTGGTAG	AAAAGAAATT	CCCATATGTT	AAGGCAGGTT	TTATGCATAT	TCCTTATATG	660
ATGGAACAGG	TGGTGAACAG	ACCGACTACT	CCAACATGA	GTTTAGTGGA	TATTCGGCGA	720
GGGATAGAAG	CAGCAATCGG	CGCTATGATA	GAACATGGAG	ATCAGGAACT	CAAGTTGGTA	780
GGCGGAGAAA	TTCATTGATA	GAAAAAAGCT	TGAGGGGAAA	ACCTTCAAGC	TTTTGGACGT	840
TTTCGAGCCA	ATACTGCTCG	GTAAACATA	ATTTTAGTGC	ATTGGATATA	AGGTAGGAGT	900
GAAAAACTAG	CAATGCCAAA	GGTAATCCAA	TTGAGGAAGT	ACCAAGGAAG	AAG	953

(2) INFORMATION FOR SEQ ID NO:84:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1060 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:84:

CTACTTGAAA	CAGAACTGAA	ATTATACCCA	CTACCTCCCT	GATTATCTTC	AATGCTTACG	60
TCTAAATAAA	CTTCCCCACT	ATTATTTAGC	TTAGCAACAA	CTGTTATAGT	AAAATAACAT	120
AAAATTCACA	TAAATAGATT	AGGGAAATCA	AAGCAACTTC	TAGGAATGTT	TTAGCAGTCA	180
CAGTGTA CTT	TCCCAGCATC	AAGCCACTAT	AACTCTGCAC	ATAAAAATGG	AGAAGATGGC	240
CATCCTCTTC	TCCAAATATT	AACTTCTTTA	CAAACCAACT	ATAGTTGACA	AAGAACCTAA	300
AATCAATTGA	TAACACGAGG	TCAGGTCGGT	CAACTCTTTC	AACTGAAGCC	CTGTCAACTC	360
TTCCCATTTA	TCAATCTTGT	ATTGGAGAGA	ATTGCGGTGC	AGATAGAGTT	GCTGGGCTGT	420
TTAAGTGAGA	ACAGCACTAT	TTTCCCAAAG	AGAGAGAATG	ATTTCCTGAA	TCTGATCTTG	480
ATCCAAAATC	ATCTGGTGTA	GACATTCTTC	GATTGGCTTC	AAGTCCACGA	GTCTTTCTCC	540
CAGACTCCAA	AGATAGAGCT	GAGAAAAAGT	ATGAACACCT	TGGTGACCCCT	GACGCCACCA	600

TGTCTTGAAC	AAATCCCGCT	CAGCTTTGAT	TAAGTCTGAT	AGGGCTTGAT	GTCCCGTCTG	660
AGACCAAACC	TGACCCAACA	TGATAGAAAG	ACGAAGTCCA	AAGTCATACT	CAACCGCTTC	720
AATCGTATCA	CTTAAAATAT	CTCTTACAGA	AGTGTATTTG	TCTTGTTGAA	GCACGAAAAC	780
ATAATCCTGA	GATCCGACCT	GTAGCACTGT	CTGACAATTC	GGAAAAAGAG	TCCGCATCAT	840
ATCTAGCCAA	GAAGCCAGAT	TTTCCTGCTG	AAAATAAGAA	AGATGGCAAT	AAACCAACTG	900
AATCTTTTTA	AAAACCTGCG	GTGCCTGTCC	CTTGCCTTCA	ACCAGATAGG	AATACCAAGG	960
GTTTAGCGAA	CGAACCTGCT	CCTGCTGGGT	CAAAAGGGCA	ACCAACTGCT	TTTCACGCTC	1020
GCTGAGCCCA	GCTTCCTCCA	GCAAAATCCA	CTGCTGAGAG			1060

(2) INFORMATION FOR SEQ ID NO:85:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 895 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:85:

ATTTTAGACT	TTGATGACAA	TCCTCAGGCG	GTTATCATGC	CCAATCACGA	GGGGCTGGAA	60
TTGCAGTTGC	CAAAGAAGTG	TGTTTATGCA	TTTTTAGGTG	AGGAGATCTG	ACCGCTATGC	120
AAGGGAAGTA	GGGGCGGATT	GTGTCGGCGA	ATTTCGTTTCT	GCTACCAAGA	CCTATCCAGT	180
CTCTTTCATC	AACTACAAGG	GTGAGGAGGT	CTGTCCTGGAT	CAGGCTCCTG	CTGGCTCCGC	240
TCCAGCAGCC	CAGTTTATGG	ATGGGTTGAT	TGGCTATGGT	GTGGAGCAGC	TTATCTCTAC	300
TGGGACCTGT	GGTGTCTAG	CTGATATAGA	GGAAAATGCC	TTTCTAGTCC	CTGTTCGCGC	360
TTTGCGAGAT	GAGGGAGCCA	GTTACCACTA	TGTGGCACCT	TGTCGTTATA	TGGAAATGCA	420
GCCAGAGGCT	ATTGCTGCTA	TTGAGGAAGT	TTTGGAAGAC	AGAGGGATTTC	CTTATGAAGA	480
AGTCATGACC	TGGACGACAG	ACGGTTTTTA	CCGAGAAACG	GCTGAAAAGG	TGGCTTATCG	540
TAAGGAAGAA	GGCTGTGCTG	TTGTGGAGAT	GGAGTGTTC	GCTCTTGCGG	CAGTAGCTCA	600
ATTGCGTG	GTTCTCTGGG	GTGAATTGTT	GTTACAGCA	AATTCTCTAG	CGGACTTGGA	660
CCAGTACAAC	AGTCGTGACT	GGGGCTCGGA	ACCTTTTAAT	AAGGCGCTAA	AACTGAGTTT	720
AGCAAGTGTC	CACCACCTTT	AGTTGTACTG	GCAAAGGATT	TGTTTTATCA	TAAAATGTCT	780
AGCTCATACT	TTTCAAAAAT	ATGTTTAAAC	GAAGTCACCT	TCCTCTTGTC	CTAAGCATGT	840
TTGAAGTTGG	GAAAAATCTT	TAAAATCAGA	AAAACGTATC	ATATCAGGTT	GATGA	895

(2) INFORMATION FOR SEQ ID NO:86:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 645 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:86:

AGGGCTGTCA	AGCTTGTTA	GAACGTTTAG	AAAAGGAGAG	TTAAGGTGGA	AAATCTTACG	60
AATTTTACG	AAAAGTATCG	TGTCTATCTG	ACTCGTCCAC	GTTTAGAGCT	TTTGGCAGTA	120
GTTACCATTG	TTTTANGNGC	TGTACTCGTC	TTTTTCTAA	ATATTCCAGG	AAAAGGTGTC	180
TTAAAACTCG	ATAATGGAAC	GATTGTTTAT	GATGGCAGTC	TTGTCCGTGG	TAAAATGAAT	240
GGCCAAGGTA	CCATTACCTT	CCAAAATGGA	GACCAATATA	CAGGTGGCTT	CAACAATGGA	300
GCCTTCAACG	GAAAAGGTAC	CTTTCAATCT	AAAGAAGGCT	GGACCTACGA	AGGTGATTTT	360
GTAAATGGTC	AGGCTGAAGG	AAAAGGGAAA	CTAACAACAG	AACAAGAAGT	CGTTTATGAA	420
GGAACTTTTA	AACAAGGCGT	TTTTCAACAA	AAATAAAGCC	TCCTTATCAA	AGGAGGTATT	480
ATTAGAATTA	CAAGGTAAGC	GTTTACCTGT	AAATCCCTTT	CTTTCCAAAT	CCCTCTTCCA	540
AGCAAGTTTG	TGAAATAAAA	AATATTTGAA	ATAAATTTCA	CAAACCTCAA	AGATAAAACC	600
TGATAAGAAA	AGAAAATGAG	AAAAGTTTCG	CAAGAGTTTA	AAAAAT		645

(2) INFORMATION FOR SEQ ID NO:87:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 572 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:87:

GAGATCTGTC	TTGACACCAA	AAGTGTGGAG	TACGCCAGCT	AATTCAACGG	CGATATAACC	60
AGCGCCTAGA	ATCGCAATTG	ACTCTGGAAG	TTCTTCCCAG	GCAAATACAT	CATCAGAAGA	120
GCCACCTAGC	TCAGCACCAG	GAATATTAGG	AATACTTGGA	TGGGCACCTG	TAGCAATCAC	180
GATATGTCTA	GCACGAATCA	GTTCAACATT	TACGCTTACA	GTATGAGAAT	CTACAAATTC	240
AGCATGACCT	TCAATCAAGT	CTACACCGTT	GCGTTTAAAA	CTACCATCAT	AGAGAAGAAC	300
GAGCGCGATC	AATGTAGGCT	TCACGATTGC	GACGTAGGGT	TGCAAAGTTA	AAGTTAAGAT	360
CAGTAGTCTC	AAAGCCGTAG	TCTCCTCCAA	ATTGATGGAA	AGTCTCAGCG	ATTTGCGCCC	420
CGCTACCACA	TGATTCTTTT	AGGAACACAA	CCGACGTTGA	CACAGGTTCC	ACCTAATTTT	480
TTTTCTCTCA	TAACGGCTGC	TTTGGCTCCA	TGTTCCCAGC	ACGGTTCATG	GTAGCGATCC	540
TCCGCTACCT	CCACGATAGC	AATGATATCA	TA			572

(2) INFORMATION FOR SEQ ID NO:88:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 49 amino acids

- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:88:

```

Val Gly Asp Asp Thr Trp Leu Phe Asp Pro Ala Lys Asp Pro Val Ile
 1           5           10           15
Met Ile Leu Pro Glu Thr Phe Phe Leu His Ala Phe Leu Leu Phe Phe
          20           25           30
Ala Leu Tyr Glu Asn Phe Phe Gly Tyr Leu Tyr Leu Lys Ser Arg Arg
          35           40           45
Lys

```

(2) INFORMATION FOR SEQ ID NO:89:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 47 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:89:

```

Val Gln Asp Phe Tyr Thr Ser Ile Asp Val Leu Ala Glu Leu Asp Asn
 1           5           10           15
Gly Thr Gln Val Ile Ile Glu Ile Gln Val His His Gln Asn Phe Ser
          20           25           30
Ser Ile Thr Cys Gly Leu Thr Cys Ala Val Arg Leu Ile Lys Ser
          35           40           45

```

(2) INFORMATION FOR SEQ ID NO:90:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 67 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:90:

```

Val Phe Ala Tyr Phe Thr Lys Pro Leu Gly Ile Lys Leu Pro Pro Tyr
 1             5             10             15
Phe Asp Ile Val His Phe Asp Gln Ala Ala Ala Ile Phe Asn Lys Tyr
             20             25             30
Pro Leu Lys Phe Val Asn Cys Val Asn Ser Ile Gly Asn Gly Leu Tyr
             35             40             45
Ile Glu Asp Glu Ser Val Val Ile Arg Pro Lys Asn Gly Phe Gly Gly
             50             55             60
Ile Gly Gly
65

```

(2) INFORMATION FOR SEQ ID NO:91:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 97 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:91:

```

Val Glu Glu Val Glu Val Ala Glu Val Lys Asn Ala Arg Val Ser Leu
 1             5             10             15
Thr Gly Glu Lys Thr Lys Pro Met Lys Leu Ala Glu Val Thr Ser Ile
             20             25             30
Asn Val Asn Arg Thr Lys Thr Glu Met Glu Glu Phe Asn Arg Val Leu
             35             40             45
Gly Gly Gly Val Val Pro Gly Lys Ser Arg Pro His Arg Trp Gly Ser
             50             55             60
Trp Asp Trp Glu Ile Asn Ser Ser Pro Thr Ser Leu Asn Pro Val Val
65             70             75             80
Pro Ser Gly Asp Ser Ser Leu Cys Gln Trp Gly Gly Val Cys Pro Ala
             85             90             95
Asp

```

(2) INFORMATION FOR SEQ ID NO:92:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 75 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:92:

```

Val Asp Val Phe Tyr Asp Gly Gln Thr Phe Thr Ile Leu Glu Asn Pro
 1             5             10             15
Val Ile Gln Gly Gln Asn Ala Gly Ala Gly Cys Thr Phe Ala Ser Ser
             20             25             30
Ile Ala Ser His Leu Val Lys Gly Asp Lys Leu Leu Pro Ala Val Glu
             35             40             45
Ser Ser Lys Ala Phe Val Tyr Arg Ala Ile Ala Gln Ala Asp Gln Tyr
             50             55             60
Gly Val Arg Gln Tyr Glu Ala Asn Lys Asn Asn
65             70             75

```

(2) INFORMATION FOR SEQ ID NO:93:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 65 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:93:

```

Val Ile Ser Val Arg Glu Lys Ser Leu Lys Val Pro Ala Ile Leu Glu
 1             5             10             15
Ala Val Glu Ala Thr Leu Gly Arg Pro Ala Phe Val Ser Phe Asp Ala
             20             25             30
Glu Lys Leu Glu Gly Ser Leu Thr Arg Leu Pro Glu Arg Asp Glu Ile
             35             40             45
Asn Pro Glu Ile Asn Glu Ala Leu Val Val Glu Phe Tyr Asn Lys Met
50             55             60

```


Leu
65

(2) INFORMATION FOR SEQ ID NO:94:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 134 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:94:

Val	Ile	Val	Glu	Lys	Glu	Glu	Lys	Gly	Glu	Glu	Met	Lys	Pro	Val	Ile
1				5				10						15	
Ser	Ile	Ile	Met	Gly	Ser	Lys	Ser	Asp	Trp	Ala	Thr	Met	Gln	Lys	Thr
			20					25					30		
Ala	Glu	Val	Leu	Asp	Arg	Phe	Gly	Val	Ala	Tyr	Glu	Lys	Lys	Val	Val
		35					40					45			
Ser	Ala	His	Arg	Thr	Pro	Asp	Leu	Met	Phe	Lys	His	Ala	Glu	Glu	Ala
	50					55					60				
Arg	Ser	Arg	Gly	Ile	Lys	Ile	Ile	Ile	Ala	Gly	Ala	Gly	Gly	Ala	Ala
65					70					75				80	
His	Leu	Pro	Gly	Met	Val	Ala	Ala	Lys	Thr	Thr	Leu	Pro	Val	Ile	Gly
				85					90					95	
Val	Pro	Val	Lys	Ser	Arg	Ala	Leu	Ser	Gly	Val	Asp	Ser	Leu	Tyr	Ser
			100					105					110		
Ile	Val	Gln	Met	Pro	Gly	Gly	Val	Pro	Val	Ala	Thr	Met	Ala	Ile	Gly
		115					120					125			
Glu	Leu	Phe	Phe	Arg	Ile										

130

(2) INFORMATION FOR SEQ ID NO:95:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 66 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

199

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:95:

Val	Arg	Xaa	Xaa	Ala	Pro	Ser	Thr	Cys	Xaa	Trp	Val	Gly	His	Met	Ala
1				5					10					15	
Ser	Gly	Leu	Arg	His	Asp	Thr	Lys	Ala	Pro	Tyr	Ser	Asp	Ser	Xaa	Xaa
			20					25					30		
Leu	Gly	Leu	Arg	Leu	Phe	Asn	Leu	Thr	Thr	Gln	Gln	Asn	Xaa	Thr	Arg
			35				40					45			
Arg	Phe	Ile	Leu	Gln	Lys	Ala	Xaa	Ser	His	Pro	Leu	Thr	Gly	Ser	Asn
	50					55					60				
Leu	Leu														
65															

(2) INFORMATION FOR SEQ ID NO:96:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 46 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:96:

Val	Asp	Asp	Thr	Asn	Thr	Leu	Asn	Val	His	Ile	His	Ala	Leu	Arg	Gln
1				5				10					15		
Glu	Leu	Ala	Lys	Tyr	Ser	Ser	Asp	Gln	Thr	Pro	Thr	Ile	Lys	Thr	Val
			20					25					30		
Trp	Gly	Leu	Gly	Tyr	Lys	Ile	Glu	Lys	Pro	Arg	Gly	Gln	Thr		
			35				40					45			

(2) INFORMATION FOR SEQ ID NO:97:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 169 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:97:

```

Val Ile Tyr Asn Ile Pro Gln Leu Ala Gly Val Ala Leu Thr Pro Ser
 1             5             10             15
Leu Tyr Thr Glu Met Leu Lys Asn Pro Arg Val Ile Gly Val Lys Asn
      20             25             30
Ser Ser Met Pro Val Gln Asp Ile Gln Thr Phe Val Ser Leu Gly Gly
      35             40             45
Glu Asp His Ile Val Phe Asn Gly Pro Asp Glu Gln Phe Leu Gly Gly
 50             55             60
Arg Leu Met Gly Ala Arg Ala Gly Ile Gly Gly Thr Tyr Gly Ala Met
65             70             75             80
Pro Glu Leu Phe Leu Lys Leu Asn Gln Leu Ile Ala Asp Lys Asp Leu
      85             90             95
Glu Thr Ala Arg Glu Leu Gln Tyr Ala Ile Asn Ala Ile Ile Gly Lys
      100            105            110
Leu Thr Ser Ala His Gly Asn Met Tyr Gly Val Ile Lys Glu Val Leu
      115            120            125
Lys Ile Asn Glu Gly Leu Asn Ile Gly Ser Val Arg Ser Pro Leu Thr
      130            135            140
Pro Val Thr Glu Glu Asp Arg Pro Val Val Glu Ala Ala Ala Ala Leu
145            150            155            160
Ile Arg Glu Thr Lys Glu Arg Phe Leu
      165

```

(2) INFORMATION FOR SEQ ID NO:98:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 288 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:98:

```

Val Thr Tyr Asp Thr Ile Gln Phe Lys Val Leu Lys Ala Val Ile Asp
 1             5             10             15
Gln Ala Phe Leu Arg Val Lys Gly Tyr Thr Leu Asn Gly His Thr Leu
      20             25             30
Pro Gly Gln Val Gln Gln Phe Asn Gln Val Phe Ile Asn Asn His Arg
      35             40             45

```

201

Ile	Thr	Pro	Glu	Val	Thr	Tyr	Lys	Lys	Ile	Asn	Glu	Thr	Thr	Ala	Glu	50	55	60
Tyr	Leu	Met	Lys	Leu	Arg	Asp	Asp	Ala	His	Leu	Ile	Asn	Ala	Glu	Met	65	70	75
Thr	Val	Arg	Leu	Gln	Val	Val	Asp	Asn	Gln	Leu	His	Phe	Asp	Val	Thr	85	90	95
Lys	Ile	Val	Asn	His	Asn	Gln	Val	Thr	Pro	Gly	Gln	Lys	Ile	Asp	Asp	100	105	110
Glu	Arg	Lys	Leu	Leu	Ser	Ser	Ile	Ser	Phe	Leu	Gly	Asn	Ala	Leu	Val	115	120	125
Ser	Val	Ser	Ser	Asp	Gln	Thr	Gly	Ala	Lys	Phe	Asp	Gly	Ala	Thr	Met	130	135	140
Ser	Asn	Asn	Thr	His	Val	Ser	Gly	Asp	Asp	His	Ile	Asp	Val	Thr	Asn	145	150	155
Pro	Met	Lys	Asp	Leu	Ala	Lys	Gly	Tyr	Met	Tyr	Gly	Phe	Val	Ser	Thr	165	170	175
Asp	Lys	Leu	Ala	Ala	Gly	Val	Trp	Ser	Asn	Ser	Gln	Asn	Ser	Tyr	Gly	180	185	190
Gly	Gly	Ser	Asn	Asp	Trp	Thr	Arg	Leu	Thr	Ala	Tyr	Lys	Glu	Thr	Val	195	200	205
Gly	Asn	Ala	Asn	Tyr	Val	Gly	Ile	His	Ser	Ser	Glu	Trp	Gln	Trp	Glu	210	215	220
Lys	Ala	Tyr	Lys	Gly	Ile	Val	Phe	Pro	Glu	Tyr	Thr	Lys	Glu	Leu	Pro	225	230	235
Ser	Ala	Lys	Val	Val	Ile	Thr	Glu	Asp	Ala	Asn	Ala	Asp	Lys	Lys	Val	245	250	255
Asp	Trp	Gln	Asp	Gly	Ala	Ile	Ala	Tyr	Arg	Ser	Ile	Met	Asn	Asn	Pro	260	265	270
Gln	Gly	Trp	Glu	Lys	Val	Lys	Asp	Ile	Thr	Ala	Met	Thr	Leu	Val	Thr	275	280	285

(2) INFORMATION FOR SEQ ID NO:99:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 66 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:99:

```

Val Ile Leu Glu Gly Asn Tyr Arg Ala Thr Ala Gly Arg Glu Glu Met
 1             5             10             15
Lys Glu Ala Ile Leu Glu Tyr Gln Ala Asn Pro Ala Ala Leu Lys Asp
             20             25             30
Leu Lys Glu Lys Ala Lys Asn Ile Ser Arg Glu Tyr Ser Glu Glu His
             35             40             45
Leu Leu Gln Ile Trp Leu Asp Phe Tyr Glu Lys Gln Ala Ala Leu Gly
             50             55             60
Thr Lys
65

```

(2) INFORMATION FOR SEQ ID NO:100:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 107 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:100:

```

Val Thr Phe Leu Asp Asp Tyr His Lys Lys His Asn Tyr Pro Leu Phe
 1             5             10             15
Tyr Glu Ser Tyr Leu Gln Asn Val Met Glu Phe Leu Glu Ser Gln Asp
             20             25             30
Ile Lys Asn Gly Val Asp Ala Phe Val Asp Asp His Gln Asn Leu Val
             35             40             45
Phe Val Leu Tyr Gly Gln Gly Tyr Arg Ala Glu Gly Lys Glu Gly Ile
             50             55             60
Leu Thr Thr Gln Val Thr Val Lys Ala Tyr Asp Glu Asp Lys Lys Pro
65             70             75             80
Ile Asn Phe Ala Asn Leu Leu Asp Ser Leu Ile Val Ser Glu Tyr Gln
             85             90             95
Met Glu Pro Asn Leu Trp Glu Val Ser Tyr Asp
             100             105

```

(2) INFORMATION FOR SEQ ID NO:101:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 185 amino acids
- (B) TYPE: amino acid

203

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:101:

```

Val Arg Lys Ser Val Pro Arg Pro Arg Leu Arg Gln Arg Ser Leu Ser
 1           5           10           15
Lys Val Ala Arg Ser Arg Leu Lys Ile Lys Lys Leu Ser Lys Val Lys
          20           25           30
His Glu Gly Gly Val Val Ile Glu Gly Ala Ser Gly Leu Leu Val Arg
          35           40           45
Ile Ala Lys Cys Cys Asn Pro Val Pro Gly Asp Asp Ile Val Gly Tyr
          50           55           60
Ile Thr Lys Gly Arg Gly Val Ala Ile His Arg Val Asp Cys Met Asn
65           70           75           80
Leu Arg Ala Gln Glu Asn Tyr Glu Gln Arg Leu Leu Asp Val Glu Trp
          85           90           95
Glu Asp Gln Tyr Ser Ser Ser Asn Lys Glu Tyr Met Ala His Ile Asp
          100          105          110
Ile Tyr Gly Leu Asn Arg Thr Gly Leu Leu Asn Asp Val Leu Gln Val
          115          120          125
Leu Ser Asn Thr Thr Lys Asn Ile Ser Thr Val Asn Ala Gln Pro Thr
          130          135          140
Lys Asp Met Lys Phe Ala Asn Ile His Val Ser Phe Gly Ile Ala Asn
145          150          155          160
Leu Ser Thr Leu Thr Thr Val Val Asp Lys Ile Lys Ser Val Pro Glu
          165          170          175
Val Tyr Ser Val Lys Arg Thr Asn Gly
          180          185

```

(2) INFORMATION FOR SEQ ID NO:102:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 115 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:102:

204

```

Val Ile Val Phe Leu Val Tyr Leu Ile Ile Thr Val Gln Lys Leu Gly
 1             5             10             15
Arg Val Ile Asp Glu Thr Glu Lys Thr Ile Lys Thr Leu Thr Ser Asp
      20             25             30
Val Asp Val Thr Leu His His Thr Asn Glu Leu Leu Ala Lys Val Asn
      35             40             45
Val Leu Ala Asp Asp Ile Asn Val Lys Val Ala Thr Ile Asp Pro Leu
      50             55             60
Phe Ser Ala Val Ala Asp Leu Ser Leu Ser Val Ser Asp Leu Asn Asp
65             70             75             80
His Ala Arg Val Leu Ser Lys Lys Ala Ser Ser Ala Gly Ser Lys Thr
      85             90             95
Leu Lys Thr Gly Ala Ser Leu Ser Ala Leu Arg Leu Ala Ser Lys Phe
      100             105             110
Phe Lys Lys
      115

```

(2) INFORMATION FOR SEQ ID NO:103:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 106 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:103:

```

Val Thr Gly Asn Trp Gln Ile Leu Phe Gln Gly Lys Met Thr Val Phe
 1             5             10             15
Ser Trp Leu Ile Gly Pro Cys Ser Ser Asp Asn Glu Glu Ala Val Leu
      20             25             30
Glu Tyr Ala Arg Arg Leu Ser Ala Leu Gln Lys Lys Val Ala Asp Lys
      35             40             45
Ile Phe Met Val Met Arg Val Tyr Thr Ala Lys Pro Arg Thr Asn Gly
      50             55             60
Asp Gly Tyr Lys Gly Leu Val His Gln Pro Asp Thr Ser Lys Ala Pro
65             70             75             80
Thr Leu Ile Asn Gly Leu Gln Ala Val Arg Gln Leu His Tyr Arg Val
      85             90             95
Asp Tyr Arg Asp Trp Phe Asp Asn Gly Arg
      205

```

100

105

(2) INFORMATION FOR SEQ ID NO:104:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 71 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:104:

Val	Gly	Thr	Gly	Ile	Ile	Gly	Ser	Ile	Val	Ser	Tyr	Pro	Val	Met	Val
1				5					10					15	
Leu	Phe	Thr	Gly	Ser	Ala	Ala	Lys	Leu	Ser	Trp	Phe	Ile	Tyr	Thr	Pro
			20					25					30		
Arg	Phe	Phe	Gly	Ala	Thr	Leu	Ile	Gly	Thr	Ala	Ile	Ser	Phe	Ile	Ala
		35					40					45			
Phe	Arg	Phe	Leu	Ile	Lys	Gln	Glu	Phe	Phe	Lys	Lys	Val	Gln	Gly	Tyr
	50					55					60				
Phe	Phe	Ala	Glu	Arg	Ile	Glu									
65						70									

(2) INFORMATION FOR SEQ ID NO:105:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 98 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:105:

Val	Ala	Ile	Ala	Arg	Gly	Leu	Ser	Met	Asn	Pro	Asp	Ile	Met	Leu	Phe
1				5					10					15	
Asp	Glu	Pro	Asn	Ser	Ala	Leu	Asp	Pro	Glu	Met	Val	Gly	Glu	Val	Ile
			20					25					30		
Asn	Val	Met	Lys	Glu	Leu	Ala	Glu	Gln	Gly	Met	Thr	Met	Ile	Ile	Val
		35					40						45		

206


```

Thr His Glu Met Gly Phe Ala Arg Gln Val Ala Asn Arg Val Ile Phe
  50                               55                               60
Thr Ala Asp Gly Glu Phe Leu Glu Asp Gly Thr Pro Asp Gln Ile Phe
  65                               70                               75                               80
Asp Asn Pro Gln His Pro Arg Leu Lys Glu Phe Leu Asp Lys Val Leu
                               85                               90                               95
Asn Val

```

(2) INFORMATION FOR SEQ ID NO:106:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 132 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:106:

```

Val Gln Ala Val Ser Glu Ser Ala Ala Ala Pro Val Arg Ala Lys Val
  1                               5                               10                               15
Arg Pro Thr Tyr Ser Thr Asn Ala Ser Ser Tyr Pro Ile Gly Glu Cys
                               20                               25                               30
Thr Trp Gly Val Lys Thr Leu Ala Pro Trp Ala Gly Asp Tyr Trp Gly
                               35                               40                               45
Asn Gly Ala Gln Trp Ala Thr Ser Ala Ala Ala Ala Gly Phe Arg Thr
                               50                               55                               60
Gly Ser Thr Pro Gln Val Gly Ala Ile Ala Cys Trp Asn Asp Gly Gly
  65                               70                               75                               80
Tyr Gly His Val Ala Val Val Thr Ala Val Glu Ser Thr Thr Arg Ile
                               85                               90                               95
Gln Val Ser Glu Ser Asn Tyr Ala Gly Asn Arg Thr Ile Gly Asn His
                               100                               105                               110
Arg Gly Trp Phe Asn Pro Thr Thr Thr Ser Glu Gly Phe Val Thr Tyr
                               115                               120                               125
Ile Tyr Ala Asp
                               130

```

(2) INFORMATION FOR SEQ ID NO:107:

(i) SEQUENCE CHARACTERISTICS:

207

- (A) LENGTH: 86 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:107:

```

Val Ile Leu Leu Asn Ser Glu Glu Lys Val Lys Lys Glu Arg Arg Ser
 1             5             10             15
Lys Glu Arg Ile Ser Thr Thr Lys Lys Gly Phe Phe Arg Met Val Leu
          20             25             30
Arg Tyr His Leu Thr Leu Leu Gly Gln Gly Thr Gly Val Val Thr Val
          35             40             45
Leu Phe Thr Ser Ala Phe Leu Pro Tyr Leu Met Met Ile Gly Leu Ile
          50             55             60
Ser Lys Ile Arg Asp Ser Gln Ile Val Pro Asp Ile His Pro Pro Tyr
65             70             75             80
Trp Leu Pro Phe Phe Leu
          85

```

(2) INFORMATION FOR SEQ ID NO:108:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 308 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:108:

```

Val Thr Pro Leu Ser Leu Leu Cys Leu Arg Lys Cys Val Arg Asp Glu
 1             5             10             15
Asn Val Phe Leu Met Gly Glu Asp Val Gly Val Phe Gly Gly Asp Phe
          20             25             30
Gly Thr Ser Val Gly Met Leu Glu Glu Phe Gly Pro Glu Arg Val Arg
          35             40             45
Asp Cys Pro Ile Ser Glu Ala Ala Ile Ser Gly Ala Ala Ala Gly Ala
          50             55             60
Ala Met Thr Gly Leu Arg Pro Ile Val Asp Met Thr Phe Met Asp Phe
          208

```

65		70		75		80									
Ser	Val	Ile	Ala	Met	Asp	Asn	Ile	Val	Asn	Gln	Ala	Ala	Lys	Thr	Arg
				85					90					95	
Tyr	Met	Phe	Gly	Gly	Lys	Gly	Gln	Val	Pro	Met	Thr	Val	Arg	Cys	Ala
				100					105					110	
Ala	Gly	Asn	Gly	Val	Gly	Ser	Ala	Ala	Gln	His	Ser	Gln	Ser	Leu	Glu
				115					120					125	
Ser	Trp	Phe	Thr	His	Ile	Pro	Gly	Leu	Lys	Val	Val	Ala	Pro	Gly	Thr
				130					135					140	
Pro	Ala	Asp	Met	Lys	Gly	Leu	Leu	Lys	Ser	Ser	Ile	Arg	Asp	Asn	Asn
				145					150					155	
Pro	Val	Ile	Ile	Leu	Glu	Tyr	Lys	Ser	Glu	Phe	Asn	Gln	Lys	Gly	Glu
				165					170					175	
Val	Pro	Val	Asp	Pro	Asp	Tyr	Thr	Ile	Pro	Leu	Gly	Val	Gly	Glu	Ile
				180					185					190	
Lys	Arg	Gln	Gly	Thr	Asp	Val	Thr	Val	Val	Thr	Tyr	Gly	Lys	Met	Leu
				195					200					205	
Arg	Arg	Val	Val	Gln	Ala	Ala	Glu	Glu	Leu	Ala	Glu	Glu	Gly	Ile	Ser
				210					215					220	
Val	Glu	Ile	Val	Asp	Pro	Arg	Thr	Leu	Val	Pro	Leu	Asp	Lys	Asp	Ile
				225					230					235	
Ile	Ile	Asn	Ser	Val	Lys	Lys	Thr	Gly	Lys	Val	Val	Leu	Val	Asn	Asp
				245					250					255	
Ala	His	Lys	Thr	Ser	Gly	Tyr	Ile	Gly	Glu	Ile	Ser	Ala	Ile	Ile	Ser
				260					265					270	
Glu	Ser	Glu	Ala	Phe	Asp	Tyr	Leu	Asp	Ala	Pro	Ile	Arg	Arg	Cys	Ala
				275					280					285	
Gly	Glu	Asp	Val	Pro	Met	Pro	Tyr	Ala	Gln	Asn	Leu	Lys	Met	Cys	Asn
				290					295					300	
Asp	Ser	Asn	Ser												
305															

(2) INFORMATION FOR SEQ ID NO:109:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 191 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:109:

209

```

Val Asp Gly Ala Thr Thr Ile Asp Ile Gly Ala Ser Thr Gly Gly Phe
 1           5           10           15
Thr Asp Val Met Leu Gln Asn Ser Ala Lys Leu Val Phe Ala Val Asp
          20           25           30
Val Gly Thr Asn Gln Leu Ala Trp Lys Leu Arg Gln Asp Pro Arg Val
          35           40           45
Val Ser Met Glu Gln Phe Asn Phe Arg Tyr Ala Glu Lys Thr Asp Phe
          50           55           60
Glu Gln Glu Pro Ser Phe Ala Ser Ile Asp Val Ser Phe Ile Ser Leu
65           70           75           80
Ser Leu Ile Leu Pro Ala Leu His Arg Val Leu Ala Asp Gln Gly Gln
          85           90           95
Val Val Ala Leu Val Lys Pro Gln Phe Glu Ala Gly Arg Glu Gln Ile
          100          105          110
Gly Lys Asn Gly Ile Ile Arg Asp Ala Lys Ile His Gln Asn Val Leu
          115          120          125
Glu Ser Val Thr Ala Met Ala Val Glu Ala Gly Phe Ser Val Leu Gly
          130          135          140
Leu Asp Phe Ser Pro Ile Gln Gly Gly His Gly Asn Ile Glu Phe Leu
145          150          155          160
Val Tyr Leu Lys Lys Glu Lys Ser Ala Ser Asn Gln Ile Leu Ala Glu
          165          170          175
Ile Lys Glu Ala Val Glu Arg Ala His Ser Gln Phe Lys Asn Glu
          180          185          190

```

(2) INFORMATION FOR SEQ ID NO:110:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 54 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:110:

```

Val Ser Ser Asp Val Lys Trp Leu Cys Gln Asn His Pro Lys Trp His
 1           5           10           15
Lys Leu Arg Gly Ile Gly Met Thr Arg Asn Thr Ile Asp Arg Asp Gly
          20           25           30
Ile Thr Ser Gln Asp Val Arg Tyr Phe Ile Phe Asn Phe Lys Leu Asp
          210

```

35 40 45
 Val Asp Asp Leu Leu Pro
 50

(2) INFORMATION FOR SEQ ID NO:111:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 126 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:111:

Val	Asp	Leu	Gln	Ser	Lys	Asn	Trp	Ser	Phe	Val	His	Arg	Phe	Ser	Glu
1				5					10					15	
Glu	Leu	Ile	Asp	Gln	His	Tyr	Gln	Asp	Leu	Val	Gly	Gln	Ser	Phe	Tyr
			20					25					30		
Pro	Pro	Ile	Arg	Glu	Phe	Met	Thr	Ser	Gly	Pro	Val	Leu	Val	Gly	Val
			35				40					45			
Ile	Ser	Gly	Pro	Lys	Val	Ile	Glu	Thr	Trp	Arg	Thr	Met	Met	Gly	Ala
	50					55				60					
Thr	Arg	Pro	Glu	Glu	Ala	Leu	Pro	Gly	Thr	Ile	Arg	Gly	Asp	Phe	Ala
65					70					75					80
Lys	Ala	Ala	Gly	Glu	Asn	Glu	Ile	Ile	Gln	Asn	Val	Val	His	Gly	Ser
			85						90				95		
Asp	Ser	Glu	Lys	Ser	Gln	Leu	Ser	Arg	Glu	Ile	Ala	Pro	Leu	Val	Leu
			100					105					110		
Arg	Val	Asp	Trp	Leu	Asn	Gln	Leu	Val	Lys	Ser	Ser	Phe	Glu		
			115				120						125		

(2) INFORMATION FOR SEQ ID NO:112:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:112:

```

Val Leu Lys Gly Val Leu Thr Leu Arg Glu Leu Thr Asn Asp Arg Asp
 1              5              10              15
Ala Asp Ile Asn Asp Phe Val Lys Val Gly Glu Val Leu Asp Val Leu
              20              25              30
Val Leu Arg Gln Val Val Gly Lys Asp Thr Asp Thr Val Thr Tyr Leu
              35              40              45
Val Ile
 50

```

(2) INFORMATION FOR SEQ ID NO:113:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 52 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:113:

```

Val Gly Glu Pro Phe Ala Asn Leu Ser Asp Leu Leu Asp Thr Tyr Tyr
 1              5              10              15
Lys Asp Lys Ala Glu Arg Asp Arg Val Lys Gln Gln Ala Ser Glu Leu
              20              25              30
Ile Arg Arg Val Glu Asn Glu Leu Gln Lys Asn Arg His Lys Leu Lys
              35              40              45
Lys Gln Glu Lys
 50

```

(2) INFORMATION FOR SEQ ID NO:114:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 113 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:114:

212

```

Val Lys Asp Lys Thr Leu Ile Ile Gln His Ser Gly Ala Tyr Ile Ala
 1              5              10              15
Arg Tyr Ser Ile Thr Trp Glu Glu Val Pro Val Asp Lys Asp Gly Asn
      20              25              30
Gln Val Val Arg Ser His Ser Trp Glu Gly Asn Gly Arg Asn Gln Thr
      35              40              45
Ala Gly Phe Val Leu Asn Leu Pro Ile Lys Glu Asn Met Arg Asn Leu
      50              55              60
Arg Val Lys Ile Glu Lys Lys Thr Gly Leu Leu Trp Asn Arg Trp Gln
65              70              75              80
Thr Ile Tyr Glu Asn Arg Pro Ile Leu Ala Gln Pro His Arg Lys Ile
      85              90              95
Thr His Trp Gly Thr Thr Leu Asn Ser Lys Val Ser Asp Asp Asp Val
      100              105              110
Leu

```

(2) INFORMATION FOR SEQ ID NO:115:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 40 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:115:

```

Val Leu Gly Ala Gly Lys Arg Leu Thr Gly Tyr Ala Ala Gly Val Glu
 1              5              10              15
Lys Lys Ala Trp Leu Leu Glu His Glu Gly Val Asp Phe Lys Asp Arg
      20              25              30
Asn Asn Arg Arg Arg Ser Thr Cys
      35              40

```

(2) INFORMATION FOR SEQ ID NO:116:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 69 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single

213

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:116:

```

Val His Val Cys Cys Ala Pro Cys Ser Thr Tyr Thr Leu Glu Tyr Leu
 1             5             10             15
Thr Lys Tyr Ala Asp Val Thr Ile Tyr Phe Ala Asn Ser Asn Ile His
      20             25             30
Pro Lys Ala Glu Tyr His Lys Arg Val Tyr Val Thr Lys Lys Phe Val
      35             40             45
Ser Asp Phe Asn Glu Gln Thr Gly Asn Thr Val Gln Tyr Leu Glu Ala
      50             55             60
Pro Tyr Glu Pro Asn
65

```

(2) INFORMATION FOR SEQ ID NO:117:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 83 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:117:

```

Val Ala Met Asp Leu Gly Phe Asp Tyr Phe Gly Ser Ala Leu Thr Ile
 1             5             10             15
Ser Pro His Lys Asn Ser Gln Thr Ile Asn Ser Ile Gly Ile Asp Val
      20             25             30
Gln Lys Ile Tyr Thr Pro His Tyr Leu Pro Asn Asp Phe Lys Lys Asn
      35             40             45
Gln Gly Tyr Lys Arg Ser Val Glu Met Arg Glu Glu Tyr Asp Ile Tyr
      50             55             60
Arg Gln Cys Tyr Cys Gly Cys Val Tyr Ala Ala Gln Ala Gln Asn Ile
      65             70             75             80
Asp Leu Val

```

(2) INFORMATION FOR SEQ ID NO:118:

214

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 52 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:118:

```

Val Thr Asp Gly Val Ile Gln Val Asp Val Leu Gly Ser Ile Val Arg
 1           5           10           15
Ser Glu Glu Trp Leu Leu Asp Asn Leu Ser Lys Gln Gly His Asp Asn
          20           25           30
Val Ala Asn Ile Phe Ile Ala Glu Tyr Asp Lys Gly Ala Val Thr Val
          35           40           45
Val Thr Tyr Lys
          50

```

(2) INFORMATION FOR SEQ ID NO:119:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 206 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:119:

```

Val Arg Glu Tyr Arg Thr Tyr Glu Glu Ile Ala Ala Asp Phe Gly Ile
 1           5           10           15
His Glu Ser Asn Leu Ile Arg Arg Ser Gln Trp Val Glu Val Thr Leu
          20           25           30
Val Gln Ser Gly Val Thr Ile Ser Lys Thr His Leu Ser Ala Glu Asn
          35           40           45
Thr Val Ile Val Asp Ala Thr Glu Val Lys Ile Asn Arg Pro Lys Lys
          50           55           60
Gln Leu Ala Asn Asp Ser Gly Lys Lys Lys Phe His Ala Met Lys Ala
        65           70           75           80
Gln Ala Ile Val Thr Ser Gln Gly Arg Ile Val Ser Leu Asp Ile Ala
          215

```

				85						90					95				
Val	Asn	Tyr	Cys	His	Asp	Met	Lys	Leu	Phe	Lys	Met	Ser	Arg	Arg	Asn				
			100						105					110					
Ile	Gly	Gln	Ala	Gly	Lys	Ile	Leu	Ala	Asp	Ser	Gly	Tyr	Gln	Gly	Pro				
		115					120					125							
Met	Lys	Ile	Tyr	Pro	Gln	Ala	Gln	Thr	Pro	Arg	Lys	Ser	Ser	Lys	Leu				
		130				135					140								
Lys	Pro	Leu	Ile	Ala	Glu	Asp	Lys	Ala	Tyr	Asn	His	Ala	Leu	Ser	Lys				
145					150				155					160					
Glu	Arg	Ser	Lys	Val	Glu	Asn	Ile	Phe	Ala	Lys	Val	Lys	Thr	Phe	Lys				
			165					170				175							
Met	Phe	Ser	Thr	Thr	Tyr	Arg	Asn	His	Arg	Lys	Arg	Phe	Gly	Leu	Arg				
			180					185				190							
Met	Asn	Leu	Ile	Ala	Gly	Ile	Ile	Asn	Tyr	Glu	Leu	Gly	Phe						
		195					200					205							

(2) INFORMATION FOR SEQ ID NO:120:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 91 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:120:

Val	Met	Gly	Pro	Gln	Gly	Asn	Gly	Phe	Asp	Leu	Ser	Asp	Leu	Asp	Glu				
1				5				10					15						
Gln	Asn	Gln	Val	Leu	Leu	Val	Gly	Gly	Gly	Ile	Gly	Val	Pro	Pro	Leu				
			20				25					30							
Leu	Glu	Val	Ala	Lys	Glu	Leu	His	Glu	Arg	Gly	Val	Lys	Val	Val	Thr				
		35				40					45								
Val	Leu	Gly	Phe	Ala	Asn	Lys	Asp	Ala	Val	Ile	Leu	Lys	Thr	Glu	Leu				
	50				55					60									
Ala	Gln	Tyr	Gly	Gln	Val	Phe	Val	Thr	Thr	Asp	Asp	Gly	Ser	Tyr	Gly				
65				70				75						80					
Ile	Lys	Gly	Asn	Val	Pro	Leu	Leu	Ser	Met	Ile									
			85					90											

(2) INFORMATION FOR SEQ ID NO:121:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 222 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:121:

```

Val Lys Met Val Leu Phe Ser Ala Gln Glu Gln Leu Tyr Tyr Lys Glu
 1              5              10              15
Lys Ile Met Thr Thr Asn Arg Leu Gln Val Ser Leu Pro Gly Leu Asp
              20              25              30
Leu Lys Asn Pro Ile Ile Pro Ala Ser Gly Cys Phe Gly Phe Gly Gln
              35              40              45
Glu Tyr Ala Lys Tyr Tyr Asp Leu Asp Leu Leu Gly Ser Ile Met Ile
              50              55              60
Lys Ala Thr Thr Leu Glu Pro Arg Phe Gly Asn Pro Thr Pro Arg Val
65              70              75              80
Ala Glu Thr Pro Ala Gly Met Leu Asn Ala Ile Gly Leu Gln Asn Pro
              85              90              95
Gly Leu Glu Val Val Leu Ala Glu Lys Leu Pro Trp Leu Glu Arg Glu
              100             105             110
Tyr Pro Asn Leu Pro Ile Ile Ala Asn Val Ala Gly Phe Ser Lys Gln
              115             120             125
Glu Tyr Ala Ala Val Ser His Gly Ile Ser Lys Ala Thr Asn Ile Lys
              130             135             140
Ala Ile Glu Leu Asn Ile Ser Cys Pro Asn Val Asp His Cys Asn His
145             150             155             160
Gly Leu Leu Ile Gly Gln Asp Pro Asp Leu Ala Tyr Asp Val Val Lys
              165             170             175
Ala Ala Val Glu Ala Ser Glu Val Pro Val Tyr Val Lys Leu Thr Pro
              180             185             190
Ser Val Thr Asp Ile Val Thr Val Ala Lys Ala Ala Glu Asp Ala Gly
              195             200             205
Ala Ser Gly Leu Thr Met Ile Ile Leu Trp Trp Asp Ala Leu
210             215             220

```

(2) INFORMATION FOR SEQ ID NO:122:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 155 amino acids

217

- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:122:

```

Val Ala Thr Gly Gln Asp Lys Ala His Ser Ile Leu Ala Ser Asn Glu
 1           5           10           15
Gly Thr Leu His Tyr Leu Val Pro Leu Lys Gln Gly Met Ser Ile Gln
          20           25           30
Gln Gly Gln Thr Ile Ala Glu Val Ser Gly Lys Glu Lys Gly Tyr Tyr
          35           40           45
Val Glu Ala Phe Val Leu Ala Ser Asp Ile Ser Arg Val Ser Lys Gly
          50           55           60
Ala Lys Val Asp Val Ala Ile Thr Gly Val Asn Ser Gln Lys Tyr Gly
65           70           75           80
Thr Leu Lys Gly Gln Val Arg Gln Ile Asp Ser Gly Thr Ile Ser Gln
          85           90           95
Glu Thr Lys Glu Gly Asn Ile Ser Leu Tyr Lys Val Met Ile Glu Leu
          100          105          110
Glu Thr Leu Thr Leu Lys His Gly Ser Glu Thr Val Ile Leu Gln Lys
          115          120          125
Asp Met Pro Val Glu Val Arg Ile Val Tyr Asp Lys Glu Thr Tyr Leu
          130          135          140
Asp Trp Ile Leu Glu Met Leu Ser Phe Lys Gln
145           150           155

```

(2) INFORMATION FOR SEQ ID NO:123:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 219 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:123:

```

Val Arg Val Pro Glu Thr Ile Thr Gln Glu Glu Leu Leu Asp Leu Ile
 1           5           10           15

```

218

```

Ala Lys Tyr Asn Gln Asp Pro Ala Trp His Gly Ile Leu Val Gln Leu
      20                      25                      30
Pro Leu Pro Lys His Ile Asp Glu Glu Ala Val Leu Leu Ala Ile Asp
      35                      40                      45
Pro Glu Lys Asp Val Asp Gly Phe His Pro Leu Asn Met Gly Arg Leu
      50                      55                      60
Trp Ser Gly His Pro Val Met Ile Pro Ser Thr Pro Ala Gly Ile Met
      65                      70                      75                      80
Glu Met Phe His Glu Tyr Gly Ile Asp Leu Glu Gly Lys Asn Ala Val
      85                      90                      95
Val Ile Gly Arg Ser Asn Ile Val Gly Lys Pro Met Ala Gln Leu Leu
      100                     105                     110
Leu Ala Lys Asn Ala Thr Val Thr Leu Ala His Ser Arg Thr His Asn
      115                     120                     125
Leu Ala Lys Val Ala Ala Lys Ala Asp Ile Leu Val Val Ala Ile Gly
      130                     135                     140
Arg Ala Lys Phe Val Thr Ala Asp Phe Val Lys Pro Gly Ala Val Val
      145                     150                     155                     160
Ile Asp Val Gly Met Asn Arg Asp Glu Asn Gly Lys Leu Cys Gly Asp
      165                     170                     175
Val Asp Tyr Glu Ala Val Ala Pro Leu Ala Ser His Ile Thr Pro Val
      180                     185                     190
Pro Gly Gly Val Gly Pro Met Thr Ile Thr Met Leu Met Glu Gln Thr
      195                     200                     205
Tyr Gln Ala Ala Leu Arg Thr Leu Asp Arg Lys
      210                     215

```

(2) INFORMATION FOR SEQ ID NO:124:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 83 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:124:

```

Val Gly Val Tyr Leu Ser Glu Gly Leu Pro Asp Leu Ile Arg Val Thr
  1                      5                      10                      15
Thr Val Thr Leu Ile Ser Leu Val Gly Glu Thr Ala Met Ala Gly Ala
      20                      25                      30
                        219

```

```

Val Gly Ala Gly Gly Ile Gly Asn Val Ala Ile Ala Tyr Gly Phe Asn
    35                                40                                45
Arg Tyr Asn His Asp Val Thr Ile Leu Ala Thr Ile Val Ile Ile Leu
    50                                55                                60
Ile Ile Phe Ala Ile Gln Phe Leu Gly Asp Phe Leu Thr Lys Lys Leu
    65                                70                                75                                80
Ser His Lys

```

(2) INFORMATION FOR SEQ ID NO:125:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 223 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:125:

```

Val Leu Pro Leu Tyr Leu Leu Phe Val Pro Tyr Gly Lys Ser Lys Lys
  1              5              10              15
Glu Val Lys Lys Arg Ala Lys Glu Ala Ser Arg Leu Thr Arg Glu Met
    20              25              30
Lys Gly Leu Ile Phe Thr Leu Ala Ile Glu Ala Ala Val Val Val Cys
    35              40              45
Thr Asn Thr Ala Ile Thr Ile Arg Ile Pro Ser Leu Met Val Glu Arg
    50              55              60
Gly Leu Gly Asp Ala Gln Leu Ser Ser Phe Val Leu Ser Ile Met Gln
    65              70              75              80
Leu Ile Gly Ile Val Ala Gly Val Ser Phe Ser Phe Leu Ile Ser Ile
    85              90              95
Phe Lys Glu Lys Leu Leu Leu Trp Ser Gly Ile Thr Phe Gly Leu Gly
    100             105             110
Gln Ile Val Ile Ala Leu Ser Ser Ser Leu Trp Val Val Val Ala Gly
    115             120             125
Ser Val Leu Ala Gly Phe Ala Tyr Ser Val Val Leu Thr Thr Val Phe
    130             135             140
Gln Leu Val Ser Glu Arg Ile Pro Ala Lys Leu Leu Asn Gln Ala Thr
    145             150             155             160
Ser Phe Ala Val Leu Gly Cys Ser Phe Gly Ala Phe Thr Thr Pro Phe
    165             170             175

```

220

Val Leu Gly Ala Ile Gly Leu Leu Thr His Asn Gly Met Leu Val Phe
 180 185 190
 Ser Ile Leu Gly Gly Trp Leu Ile Val Ile Ser Ile Phe Val Met Tyr
 195 200 205
 Leu Leu Gln Lys Arg Ala Leu Gly Leu Ile Pro Lys Phe Phe Phe
 210 215 220

(2) INFORMATION FOR SEQ ID NO:126:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 119 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:126:

Val Val Ala Gly Pro Glu Gly Leu Asp Glu Ala Gly Leu Asn Gly Thr
 1 5 10 15
 Thr Xaa Ile Ala Leu Xaa Glu Asn Gly Glu Ile Ser Leu Ser Ser Phe
 20 25 30
 Thr Pro Glu Asp Leu Gly Met Glu Gly Tyr Ala Met Glu Asp Ile Arg
 35 40 45
 Gly Gly Asn Ala Gln Glu Asn Ala Glu Ile Leu Leu Ser Val Leu Lys
 50 55 60
 Asn Glu Ala Ser Pro Phe Leu Glu Thr Thr Val Leu Asn Ala Gly Leu
 65 70 75 80
 Gly Phe Tyr Ala Asn Gly Lys Ile Asp Ser Ile Lys Glu Gly Val Ala
 85 90 95
 Leu Ala Arg Gln Val Ile Ala Arg Gly Lys Ala Leu Glu Lys Leu Arg
 100 105 110
 Leu Leu Gln Glu Tyr Gln Lys
 115

(2) INFORMATION FOR SEQ ID NO:127:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 112 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:127:

```

Val Asp Ile Val Gln Gln Ala Gln Thr Tyr Glu Glu Asn Gly Ala Val
 1             5             10             15
Met Ile Ser Val Leu Thr Asp Glu Val Phe Phe Lys Gly His Leu Asp
          20             25             30
Tyr Leu Arg Glu Ile Ser Ser Gln Val Glu Ile Pro Thr Leu Asn Lys
          35             40             45
Asp Phe Ile Ile Asp Glu Lys Gln Ile Ile Arg Ala Arg Asn Ala Gly
          50             55             60
Ala Thr Val Ile Leu Leu Ile Val Ala Ala Leu Ser Glu Glu Arg Leu
65             70             75             80
Lys Glu Leu Tyr Asp Tyr Ala Thr Glu Leu Gly Leu Glu Val Leu Val
          85             90             95
Glu Thr His Asn Leu Ala Glu Leu Glu Val Ala His Arg Leu Gly Gly
          100            105            110

```

(2) INFORMATION FOR SEQ ID NO:128:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 65 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:128:

```

Val Ser Glu Lys His Ala Gly Phe Met Ile Asn Val Ala Asp Gly Thr
 1             5             10             15
Ala Lys Asp Tyr Glu Asp Leu Ile Gln Ser Val Ile Glu Lys Val Lys
          20             25             30
Glu His Ser Gly Ile Thr Leu Glu Arg Glu Val Arg Ile Leu Gly Glu
          35             40             45
Ser Leu Ser Val Ala Lys Met Tyr Ala Gly Gly Phe Thr Pro Cys Lys
          50             55             60
Arg
65

```


(2) INFORMATION FOR SEQ ID NO:129:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 29 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:129:

```

Val Glu Arg Ile Ile Arg Lys Ala Phe Ala Ile Glu Leu Gln Glu Ile
 1             5             10             15
Ala Glu Lys Ser Leu Leu Val Ser Ile Ser Lys Met Phe
      20             25

```

(2) INFORMATION FOR SEQ ID NO:130:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 88 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:130:

```

Val Arg Ile Gly Asn Thr Val Leu Ala Asn Val Thr Ser Gly Val Ala
 1             5             10             15
Lys Gln Ala Ser Lys Ala Ala Gln Ala Ser Asn Leu Gly Gly Gly Ala
      20             25             30
Glu Val Asp Gly Phe Ser Lys Thr Leu Ser Ser Leu Asp Ile Ser Ile
      35             40             45
Gln Thr Ser Asp Phe Ile Ile Ile Phe Val Leu Ala Leu Val Leu Val
      50             55             60
Val Leu Val Met Ala Leu Ala Ser Ser Asn Leu Leu Arg Lys Gln Pro
      65             70             75             80
Lys Glu Leu Leu Leu Asp Gly Glu
      85

```

(2) INFORMATION FOR SEQ ID NO:131:

223

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 164 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:131:

Val	Ser	Asn	Lys	Thr	Phe	Pro	Ile	Leu	Val	Asn	Lys	Asp	Pro	Lys	Thr
1				5					10					15	
Gly	Thr	Tyr	Ser	Gly	Ile	Glu	Thr	Asp	Leu	Ala	Lys	Met	Val	Ala	Asp
			20					25					30		
Glu	Leu	Lys	Val	Lys	Ile	His	Tyr	Val	Pro	Val	Thr	Ala	Gln	Thr	Arg
			35				40					45			
Gly	Pro	Leu	Leu	Asp	Asn	Glu	Gln	Val	Asp	Met	Asp	Ile	Ala	Thr	Phe
	50					55				60					
Thr	Ile	Thr	Asp	Glu	Arg	Lys	Lys	Leu	Tyr	Asn	Phe	Thr	Ser	Pro	Tyr
65					70					75				80	
Tyr	Thr	Asp	Ala	Ser	Gly	Phe	Leu	Val	Asn	Lys	Ser	Ala	Lys	Ile	Lys
				85					90				95		
Lys	Ile	Glu	Asp	Leu	Asn	Gly	Lys	Thr	Ile	Gly	Val	Ala	Gln	Gly	Ser
			100					105					110		
Ile	Thr	Gln	Arg	Leu	Ile	Thr	Glu	Leu	Gly	Lys	Lys	Lys	Gly	Leu	Lys
		115					120					125			
Phe	Lys	Phe	Val	Glu	Leu	Gly	Ser	Tyr	Pro	Glu	Leu	Ile	Thr	Ser	Leu
	130					135				140					
His	Ala	His	Arg	Ile	Asp	Ala	Phe	Ser	Val	Asp	Arg	Ser	Ile	Leu	Ser
145					150					155				160	
Gly	Tyr	Thr	Ser												

(2) INFORMATION FOR SEQ ID NO:132:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 62 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:132:

```

Val Leu Glu Glu Leu Arg Ile Pro Ala Pro Asn Glu Phe Glu Asp Leu
 1             5             10             15
Asp Leu Ser Pro Leu Asp Phe Lys Pro His Ile Ala Pro His Lys Phe
          20             25             30
Glu Gly Met Val Glu Thr Ala Arg Asp Leu Ile Arg Asn Gly Asp Met
          35             40             45
Phe Arg Cys Val Thr Gln Pro Ala Phe Ser Ser Arg Arg Ser
          50             55             60

```

(2) INFORMATION FOR SEQ ID NO:133:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 65 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:133:

```

Val Ser Ser Ser Phe Phe Thr Pro Leu Lys Gln Leu Ser Lys Phe Leu
 1             5             10             15
Ile Ile Met Ala Met Ser Ala Ile Gly Leu Lys Thr Asn Leu Val Ala
          20             25             30
Met Val Lys Ser Ser Gly Lys Ser Ile Val Leu Gly Ala Val Cys Trp
          35             40             45
Ile Ala Ile Ile Leu Thr Ser Leu Gly Met Gln Thr Leu Ile Gly Ile
          50             55             60
Phe
65

```

(2) INFORMATION FOR SEQ ID NO:134:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 71 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:134:

```

Val Pro Glu Asp Tyr Arg Ile Ile Thr Ser Asp Asp Ser Gln Ile Ser
 1             5             10             15
Arg Phe Thr Arg Pro Asn Leu Thr Thr Ile Ala Gln Pro Leu Tyr Asp
          20             25             30
Leu Gly Ala Ile Ser Met Arg Met Leu Thr Lys Ile Met His Lys Glu
          35             40             45
Glu Leu Glu Glu Arg Glu Val Leu Leu Pro His Gly Leu Thr Glu Arg
          50             55             60
Ser Ser Thr Arg Lys Arg Lys
65             70

```

(2) INFORMATION FOR SEQ ID NO:135:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 163 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:135:

```

Val Gly Gln Ser Gln Phe Leu Phe Lys Val Ser Tyr Ala Asp Gly Gln
 1             5             10             15
Lys Ala Tyr Arg Val Asp Leu Pro Asp Leu Leu Thr Lys Thr Asp Trp
          20             25             30
Gln Ile Ile Lys Ser Phe Leu Asp Val Leu Leu Ala Tyr Thr Gly Thr
          35             40             45
Asp Ile Glu Gly Leu Asp Gly Phe Asp Phe Glu Ala Tyr Phe Gln Ala
          50             55             60
Ser Ile Gln Ala Tyr Leu Ala Asp Pro Val Ala Arg Phe Thr Ile Cys
65             70             75             80
Gln Arg Ile Phe Asn Pro Ile Phe Phe Ser Arg Glu Asn Leu Lys Ser
          85             90             95
Phe Leu Glu Ala Asp Gly Leu Ala Gln Phe Glu Ala Arg Val Arg Ala
          100            105            110
Val Gln Glu Thr Asp Ala Tyr Phe Ala Arg Val Ser Phe Tyr Gln Asp
          115            120            125

```

226

Gly Glu Gly Lys Val His Gly Val Tyr His Leu Ala Gln Gly Val Lys
 130 135 140
 Thr Val Leu Pro Arg Glu Pro Phe Val Pro Ala Ala Tyr Ile Glu Arg
 145 150 155 160
 Ile Gly Gly

(2) INFORMATION FOR SEQ ID NO:136:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 69 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:136:

Val Asp Lys Glu Val Gln Trp Glu Ile Asp Leu Val Gln Ile Thr Gly
 1 5 10 15
 Asp Gly Ser Lys Pro Glu Asp Tyr Glu Ser Ile Ala Arg Leu Asp Tyr
 20 25 30
 Ala Lys Phe Leu Glu Val Leu Pro Pro Ser Phe Tyr His Gln Leu Asp
 35 40 45
 Ala Asn Gln Ile Glu Ile Gln Pro Ile Leu Gly Gln Asp Phe Lys Thr
 50 55 60
 Leu Ala Gln Glu Lys
 65

(2) INFORMATION FOR SEQ ID NO:137:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 299 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:137:

Val Ile Leu Lys Ile Glu Asp Leu Val Met Ser Ile Ile Ser Thr Asp
 227

1	5	10	15
Leu Thr Pro Phe Gln Ile Asp Asp Thr Leu Lys Ala Ala Leu Arg Glu			
20	25	30	
Asp Val His Ser Glu Asp Tyr Ser Thr Asn Ala Ile Phe Asp His His			
35	40	45	
Gly Gln Ala Lys Val Ser Leu Phe Ala Lys Glu Ala Gly Val Leu Ala			
50	55	60	
Gly Leu Thr Val Phe Gln Arg Val Phe Thr Leu Phe Asp Ala Glu Val			
65	70	75	80
Thr Phe Gln Asn Pro His Gln Phe Lys Asp Gly Asp Arg Leu Thr Ser			
85	90	95	
Gly Asp Leu Val Leu Glu Ile Ile Gly Ser Val Arg Ser Leu Leu Thr			
100	105	110	
Cys Glu Arg Val Ala Leu Asn Phe Leu Gln His Leu Ser Gly Ile Ala			
115	120	125	
Ser Met Thr Ala Ala Tyr Val Glu Ala Leu Gly Asp Asp Cys Ile Lys			
130	135	140	
Val Phe Asp Thr Arg Lys Thr Thr Pro Asn Leu Arg Leu Phe Glu Lys			
145	150	155	160
Tyr Ala Val Arg Val Gly Gly Gly Tyr Asn His Arg Phe Asn Leu Ser			
165	170	175	
Asp Ala Ile Leu Leu Lys Asp Asn His Ile Ala Ala Val Gly Ser Val			
180	185	190	
Gln Arg Ala Ile Ala Gln Ala Arg Ala Tyr Ala Pro Phe Val Lys Met			
195	200	205	
Val Glu Val Glu Val Glu Ser Leu Ala Ala Ala Glu Glu Ala Ala Ala			
210	215	220	
Ala Gly Ala Asp Ile Ile Met Leu Asp Asn Met Ser Leu Glu Gln Ile			
225	230	235	240
Glu Gln Ala Ile Thr Leu Ile Ala Gly Arg Ser Arg Ile Glu Cys Ser			
245	250	255	
Gly Asn Ile Asp Met Thr Thr Ile Ser Arg Phe Arg Gly Leu Ala Ile			
260	265	270	
Asp Tyr Val Ser Ser Gly Ser Leu Thr His Ser Ala Lys Ser Leu Asp			
275	280	285	
Phe Ser Met Lys Gly Leu Thr Tyr Leu Asp Val			
290	295		

(2) INFORMATION FOR SEQ ID NO:138:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 242 amino acids.

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:138:

Val	Glu	Val	Glu	Val	Pro	Thr	Gln	Val	Pro	Ala	His	Ile	Gly	Ile	Ile	1	5	10	15
Met	Asp	Gly	Asn	Gly	Arg	Trp	Ala	Lys	Lys	Arg	Met	Gln	Pro	Arg	Val	20	25	30	
Phe	Gly	His	Lys	Ala	Gly	Met	Glu	Ala	Leu	Gln	Thr	Val	Thr	Lys	Ala	35	40	45	
Ala	Asn	Lys	Leu	Gly	Val	Lys	Val	Ile	Thr	Val	Tyr	Ala	Phe	Ser	Thr	50	55	60	
Glu	Asn	Trp	Thr	Arg	Pro	Asp	Gln	Glu	Val	Lys	Phe	Ile	Met	Asn	Leu	65	70	75	80
Pro	Val	Glu	Phe	Tyr	Asp	Asn	Tyr	Val	Pro	Glu	Leu	His	Ala	Asn	Asn	85	90	95	
Val	Lys	Ile	Gln	Met	Ile	Gly	Glu	Thr	Asp	Arg	Leu	Pro	Lys	Gln	Thr	100	105	110	
Phe	Glu	Ala	Leu	Thr	Lys	Ala	Glu	Glu	Leu	Thr	Lys	Asn	Asn	Thr	Gly	115	120	125	
Leu	Ile	Leu	Asn	Phe	Ala	Leu	Asn	Tyr	Gly	Gly	Arg	Ala	Glu	Ile	Thr	130	135	140	
Gln	Ala	Leu	Lys	Leu	Ile	Ser	Gln	Asp	Val	Leu	Asp	Ala	Lys	Ile	Asn	145	150	155	160
Pro	Gly	Asp	Ile	Thr	Glu	Glu	Leu	Ile	Gly	Asn	Tyr	Leu	Phe	Thr	Gln	165	170	175	
His	Leu	Pro	Lys	Asp	Leu	Arg	Asp	Pro	Asp	Leu	Ile	Ile	Arg	Thr	Ser	180	185	190	
Gly	Glu	Leu	Arg	Leu	Ser	Asn	Phe	Leu	Pro	Trp	Gln	Gly	Ala	Tyr	Ser	195	200	205	
Glu	Leu	Tyr	Phe	Thr	Asp	Thr	Leu	Trp	Pro	Asp	Phe	Asp	Glu	Ala	Ala	210	215	220	
Leu	Gln	Glu	Ala	Ile	Leu	Ala	Tyr	Asn	Arg	Arg	His	Arg	Arg	Phe	Gly	225	230	235	240
Gly	Val																		

(2) INFORMATION FOR SEQ ID NO:139:

(i) SEQUENCE CHARACTERISTICS:

229

- (A) LENGTH: 183 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:139:

```

Val Val Ala Tyr Ser Val Leu Ile Ser Ile Met Leu Gly Thr Thr Val
 1           5           10           15
Phe Ser Lys Ser Tyr Thr Ile Glu Asp Ala Val Phe Pro Leu Ala Met
      20           25           30
Ser Phe Tyr Val Gly Phe Gly Phe Asn Ala Leu Leu Asp Ala Arg Val
      35           40           45
Ala Gly Leu Asp Lys Ala Leu Leu Ala Leu Cys Ile Val Trp Ala Thr
      50           55           60
Asp Ser Gly Ala Tyr Leu Val Gly Met Asn Tyr Gly Lys Arg Lys Leu
65           70           75           80
Ala Pro Arg Val Ser Pro Asn Lys Thr Leu Glu Gly Ala Leu Gly Gly
      85           90           95
Ile Leu Gly Ala Ile Leu Val Thr Ile Ile Phe Met Ile Val Asp Ser
      100          105          110
Thr Val Ala Leu Pro Tyr Gly Ile Tyr Lys Met Ser Val Phe Ala Ile
      115          120          125
Phe Phe Ser Ile Ala Gly Gln Phe Gly Asp Leu Leu Glu Ser Ser Ile
      130          135          140
Lys Arg His Phe Gly Val Lys Asp Ser Gly Lys Phe Ile Pro Gly His
145          150          155          160
Gly Gly Val Leu Asp Arg Phe Asp Ser Met Leu Leu Val Phe Pro Ile
      165          170          175
Met His Leu Phe Gly Leu Phe
      180

```

(2) INFORMATION FOR SEQ ID NO:140:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 95 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

230

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:140:

Val	Asp	Leu	Leu	Leu	Ser	Leu	Arg	Gln	Val	Val	Met	Leu	Leu	Lys	Met
1				5					10					15	
Glu	Leu	Arg	Ile	Phe	Leu	Tyr	Phe	Leu	Ala	Met	Ile	Ser	Ile	Asn	Ile
			20					25					30		
Gly	Ile	Phe	Asn	Leu	Ile	Pro	Ile	Pro	Ala	Leu	Asp	Gly	Gly	Lys	Ile
		35				40					45				
Val	Leu	Asn	Ile	Leu	Glu	Ala	Ile	Arg	Arg	Lys	Pro	Leu	Lys	Gln	Glu
	50					55				60					
Ile	Glu	Thr	Tyr	Val	Thr	Leu	Ala	Gly	Val	Val	Ile	Met	Val	Val	Leu
65					70					75					80
Met	Ile	Ala	Val	Thr	Trp	Asn	Asp	Ile	Met	Arg	Leu	Phe	Phe	Arg	
			85					90						95	

(2) INFORMATION FOR SEQ ID NO:141:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 183 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:141:

Val	Glu	Leu	Met	Ser	Thr	Val	Gln	Lys	Ser	Thr	Phe	Met	Lys	Cys	Val
1				5				10					15		
Asn	Thr	Leu	Glu	Trp	Phe	Phe	Asn	Ala	Pro	Ile	His	Leu	Leu	Asn	Arg
		20					25					30			
Ile	Tyr	Arg	Asn	Ile	Thr	Phe	Ala	His	Glu	Arg	Ala	Gly	Val	Lys	Asp
	35					40					45				
Lys	Gln	Val	Leu	Asp	Glu	Ile	Val	Glu	Thr	Ser	Leu	Ser	Gln	Ala	Ala
	50					55				60					
Leu	Trp	Asp	Gln	Val	Lys	Asp	Asp	Leu	His	Lys	Ser	Ala	Leu	Thr	Leu
65				70						75					80
Ser	Gly	Gly	Gln	Gln	Gln	Arg	Leu	Cys	Ile	Ala	Arg	Ala	Ile	Ser	Val
			85					90					95		
Lys	Pro	Asp	Ile	Leu	Leu	Met	Asp	Glu	Pro	Ala	Ser	Ala	Leu	Asp	Pro
		100					105					110			
Ile	Ala	Thr	Met	Gln	Leu	Glu	Glu	Thr	Met	Phe	Glu	Leu	Lys	Lys	Asn
															231

115	120	125
Phe Thr Ile Ile Ile Val Thr His Asn Met Gln Gln Ala Ala Arg Ala		
130	135	140
Ser Asp Tyr Thr Gly Phe Phe Tyr Leu Gly Asp Leu Ile Glu Tyr Asp		
145	150	155
Lys Thr Ala Thr Ile Phe Gln Asn Ala Lys Leu Gln Ser Thr Asn Asp		
165	170	175
Tyr Val Ser Gly His Phe Gly		
180		

(2) INFORMATION FOR SEQ ID NO:142:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 228 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:142:

Val	Pro	Lys	Glu	Ser	Leu	Thr	Gln	Val	Leu	Pro	Arg	Asp	Leu	His	Ala
1				5				10						15	
Glu	Tyr	Phe	Ala	Val	Leu	Ala	Ser	Ile	Ala	Thr	Ser	Ile	Glu	Arg	Met
			20					25					30		
Ala	Thr	Glu	Ile	Arg	Gly	Leu	Gln	Lys	Ser	Glu	Gln	Arg	Glu	Val	Glu
		35					40					45			
Glu	Phe	Phe	Ala	Lys	Gly	Gln	Lys	Gly	Ser	Ser	Ala	Met	Pro	His	Lys
	50					55					60				
Arg	Asn	Pro	Ile	Gly	Ser	Glu	Asn	Met	Thr	Gly	Leu	Ala	Arg	Val	Ile
65					70				75					80	
Arg	Gly	His	Met	Ile	Thr	Ala	Tyr	Glu	Asn	Val	Ala	Leu	Trp	His	Glu
			85						90					95	
Arg	Asp	Ile	Ser	His	Ser	Ser	Ala	Glu	Arg	Ile	Ile	Thr	Pro	Asp	Thr
		100						105					110		
Thr	Ile	Leu	Ile	Asp	Tyr	Met	Leu	Asn	Arg	Phe	Gly	Asn	Ile	Val	Lys
		115					120					125			
Asn	Leu	Thr	Val	Phe	Pro	Glu	Asn	Met	Ile	Arg	Asn	Met	Asn	Ser	Thr
	130					135					140				
Phe	Gly	Leu	Ile	Phe	Ser	Gln	Arg	Ala	Met	Leu	Thr	Leu	Ile	Glu	Lys
145				150					155					160	
Gly	Met	Thr	Arg	Glu	Gln	Ala	Tyr	Asp	Leu	Val	Gln	Pro	Lys	Thr	Ala
															232

	165		170		175
Tyr Ser Trp Asp Asn Gln Val Asp Phe Lys Pro Leu Leu Glu Ala Asp					
	180		185		190
Ser Glu Val Thr Ser Arg Leu Thr Gln Glu Glu Ile Asp Glu Ile Phe					
	195		200		205
Asn Pro Val Tyr Tyr Thr Lys Arg Val Asp Asp Ile Phe Glu Arg Leu					
	210		215		220
Gly Leu Gly Asp					
225					

(2) INFORMATION FOR SEQ ID NO:143:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 157 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:143:

Val Ile Phe Ile Ser Thr Leu Ser Leu Gly Gly Leu Ala His Leu Leu			
1	5	10	15
Trp Phe Ser Leu Pro Leu Ala Ala Cys Leu Ala Val Gly Ala Ala Leu			
	20	25	30
Gly Pro Thr Asp Leu Val Ala Phe Ala Ser Leu Ser Glu Arg Phe Ser			
	35	40	45
Phe Pro Lys Arg Val Ser Asn Ile Leu Lys Gly Glu Gly Leu Leu Asn			
	50	55	60
Asp Ala Ser Gly Leu Val Ala Phe Gln Val Ala Leu Thr Ala Trp Thr			
65	70	75	80
Thr Gly Ala Phe Ser Leu Gly Gln Ala Ser Ser Ser Leu Ile Phe Ser			
	85	90	95
Ile Leu Gly Gly Phe Leu Ile Gly Phe Leu Thr Ala Met Thr Asn Arg			
	100	105	110
Phe Leu His Thr Phe Leu Leu Ser Val Arg Ala Thr Asp Ile Ala Ser			
	115	120	125
Glu Leu Leu Leu Glu Phe Glu Phe Ala Ser Ser Asp Leu Leu Ser Gly			
	130	135	140
Arg Arg Ser Pro Cys Phe Arg Asp Tyr Cys Arg Arg Ser			
145	150	155	

(2) INFORMATION FOR SEQ ID NO:144:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 230 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:144:

```

Val Thr Phe Phe Leu Ala Glu Glu Val His Val Ser Gly Ile Ile Ala
 1             5             10             15
Val Val Val Asp Arg Ile Leu Lys Ala Ser Arg Phe Lys Lys Ile Thr
      20             25             30
Leu Leu Glu Ala Gln Val Asp Thr Val Thr Glu Thr Val Trp His Thr
      35             40             45
Val Thr Phe Met Leu Asn Gly Ser Val Phe Val Ile Leu Gly Met Glu
      50             55             60
Leu Glu Met Ile Ala Glu Pro Ile Leu Thr Asn Pro Ile Tyr Asn Pro
      65             70             75             80
Leu Leu Leu Leu Leu Ser Leu Ile Ala Leu Thr Phe Val Leu Phe Val
      85             90             95
Ile Arg Phe Ile Met Ile Tyr Gly Tyr Tyr Ala Tyr Arg Thr Arg Arg
      100             105             110
Leu Lys Lys Lys Leu Asn Lys Tyr Met Lys Asp Met Phe Leu Leu Thr
      115             120             125
Phe Ser Gly Val Lys Gly Thr Val Ser Ile Ala Thr Ile Leu Leu Ile
      130             135             140
Pro Ser Asn Leu Glu Gln Glu Tyr Pro Leu Leu Leu Phe Leu Val Ala
      145             150             155             160
Gly Val Thr Leu Val Ser Phe Leu Thr Gly Leu Leu Val Leu Pro His
      165             170             175
Leu Ser Asp Glu Glu Glu Glu Ser Lys Asp Tyr Leu Met His Ile Ala
      180             185             190
Ile Leu Asn Glu Val Thr Leu Glu Leu Glu Lys Glu Leu Glu Asp Thr
      195             200             205
Arg Asn Lys Leu Pro Leu Tyr Ala Ala Ile Asp Asn Ser Ile Met Asp
      210             215             220
Val Leu Lys Ile Ser Phe
      225             230

```

(2) INFORMATION FOR SEQ ID NO:145:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 98 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:145:

```

Val Thr Gly Glu Val Gly Asp Leu Lys Gln Gly Phe Ser Val Asn Ile
 1           5           10           15
Glu Val Lys Ser Lys Thr Lys Ala Ile Leu Val Pro Val Ser Ser Leu
          20           25           30
Val Met Asp Asp Ser Lys Asn Tyr Val Trp Ile Val Asp Glu Gln Gln
          35           40           45
Lys Ala Lys Lys Val Glu Val Ser Leu Gly Asn Ala Asp Ala Glu Asn
          50           55           60
Gln Glu Ile Thr Ser Gly Leu Thr Asn Gly Ala Lys Val Ile Ser Asn
65           70           75           80
Pro Thr Ser Ser Leu Glu Glu Gly Lys Glu Val Lys Ala Asp Glu Ala
          85           90           95
Thr Asn

```

(2) INFORMATION FOR SEQ ID NO:146:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 182 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:146:

```

Val Gly Leu Gln Ile Arg Ala Ile Phe Lys Arg Tyr Thr Asp Leu Ile
 1           5           10           15
Glu Pro Met Ser Ile Asp Glu Ala Tyr Leu Asp Val Thr Glu Asn Lys
          20           25           30
                               235

```

```

Leu Gly Ile Lys Ser Ala Val Lys Ile Ala Arg Leu Ile Gln Lys Asp
   35                               40                               45
Ile Trp Gln Glu Leu His Leu Thr Ala Ser Ala Gly Val Ser Tyr Asn
   50                               55                               60
Lys Phe Leu Ala Lys Met Ala Ser Asp Tyr Gln Lys Pro His Gly Leu
  65                               70                               75                               80
Thr Val Ile Leu Pro Glu Gln Ala Glu Asp Phe Leu Lys Gln Met Asp
                               85                               90                               95
Ile Ser Lys Phe His Gly Val Gly Lys Lys Thr Val Glu Arg Leu His
   100                               105                               110
Gln Met Gly Val Phe Thr Gly Ala Asp Leu Leu Glu Val Pro Glu Val
   115                               120                               125
Thr Leu Ile Asp Arg Phe Gly Arg Leu Gly Tyr Asp Leu Tyr Arg Lys
  130                               135                               140
Ala Arg Gly Ile His Asn Ser Pro Val Lys Ser Asn His Ile Arg Lys
  145                               150                               155                               160
Ser Ile Gly Lys Glu Lys Thr Tyr Gly Lys Ile Leu Arg Ala Glu Glu
                               165                               170                               175
Asp Ile Lys Lys Glu Ser
                               180

```

(2) INFORMATION FOR SEQ ID NO:147:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 343 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:147:

```

Val Asn Leu Pro Lys Arg Ala Phe Leu Asn Gly Arg Val Asp Leu Thr
  1                               5                               10                               15
Gln Ala Glu Ala Val Met Asp Ile Ile Arg Ala Lys Thr Asp Lys Ala
   20                               25                               30
Met Asn Ile Ala Val Lys Gln Leu Asp Gly Ser Leu Ser Asp Leu Ile
   35                               40                               45
Asn Asn Thr Arg Gln Glu Ile Leu Asn Thr Leu Ala Gln Val Glu Val
   50                               55                               60
Asn Ile Asp Tyr Pro Glu Tyr Asp Asp Val Glu Glu Ala Thr Thr Ala
  65                               70                               75                               80

```

```

Val Val Arg Glu Lys Thr Met Glu Phe Glu Gln Leu Leu Thr Lys Leu
      85                      90                      95
Leu Arg Thr Ala Arg Arg Gly Lys Ile Leu Arg Glu Gly Ile Ser Thr
      100                      105                      110
Ala Ile Ile Gly Arg Pro Asn Val Gly Lys Ser Ser Leu Leu Asn Asn
      115                      120                      125
Leu Leu Arg Glu Asp Lys Ala Ile Val Thr Asp Ile Ala Gly Thr Thr
      130                      135                      140
Arg Asp Val Ile Glu Glu Tyr Val Asn Ile Asn Gly Val Pro Leu Lys
145                      150                      155                      160
Leu Ile Asp Thr Ala Gly Ile Arg Glu Thr Asp Asp Ile Val Glu Gln
      165                      170                      175
Ile Gly Val Glu Arg Ser Lys Lys Ala Leu Lys Glu Ala Asp Leu Val
      180                      185                      190
Leu Leu Val Leu Asn Ala Ser Glu Pro Leu Thr Ala Gln Asp Arg Gln
      195                      200                      205
Leu Leu Glu Ile Ser Gln Asp Thr Asn Arg Ile Ile Leu Leu Asn Lys
      210                      215                      220
Thr Asp Leu Pro Glu Thr Ile Glu Thr Ser Lys Leu Pro Glu Asp Val
225                      230                      235                      240
Ile Arg Ile Ser Val Leu Lys Asn Gln Asn Ile Asp Lys Ile Glu Glu
      245                      250                      255
Arg Ile Asn Asn Leu Phe Phe Glu Asn Ala Gly Leu Val Glu Gln Asp
      260                      265                      270
Ala Thr Tyr Leu Ser Asn Ala Arg His Ile Ser Leu Ile Glu Lys Ala
      275                      280                      285
Val Glu Ser Leu Gln Ala Val Asn Gln Gly Leu Glu Leu Gly Met Pro
      290                      295                      300
Val Asp Leu Leu Gln Val Asp Leu Thr Arg Thr Trp Glu Ile Leu Gly
305                      310                      315                      320
Glu Ile Thr Gly Asp Ala Ala Pro Asp Glu Leu Ile Thr Gln Leu Phe
      325                      330                      335
Ser Gln Phe Cys Leu Gly Lys
      340

```

(2) INFORMATION FOR SEQ ID NO:148:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 115 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:148:

```

Val Glu Ile Ser Val Gln Pro Pro Gly Lys Lys Ile Gln Ser Leu Asn
 1           5           10           15
Leu Met Ser Gly Gly Glu Lys Ala Leu Ser Ala Leu Ala Leu Leu Phe
          20           25           30
Ser Ile Ile Arg Val Lys Thr Ile Pro Phe Val Ile Leu Asp Glu Val
        35           40           45
Glu Ala Ala Leu Asp Glu Ala Asn Val Lys Arg Phe Gly Asp Tyr Leu
       50           55           60
Asn Arg Phe Asp Lys Asp Ser Gln Phe Ile Val Val Thr His Arg Lys
65           70           75           80
Gly Thr Met Ala Ala Ala Asp Ser Ile Tyr Gly Val Thr Met Gln Glu
          85           90           95
Ser Gly Val Ser Lys Ile Val Ser Val Lys Leu Lys Asp Leu Glu Ser
          100          105          110
Ile Glu Gly
        115

```

(2) INFORMATION FOR SEQ ID NO:149:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 231 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:149:

```

Val Thr Thr Val Ala Glu Phe Gly Asp Ser Ser Lys Leu Thr Val Gly
 1           5           10           15
Glu Thr Ala Ile Ala Ile Gly Ser Pro Leu Gly Ser Glu Tyr Ala Asn
          20           25           30
Thr Val Thr Gln Gly Ile Val Ser Ser Leu Asn Arg Asn Val Ser Leu
        35           40           45
Lys Ser Glu Asp Gly Gln Ala Ile Ser Thr Lys Ala Ile Gln Thr Asp
       50           55           60
Thr Ala Ile Asn Pro Gly Asn Ser Gly Gly Pro Leu Ile Asn Ile Gln
65           70           75           80

```

238


```

Gly Gln Val Ile Gly Ile Thr Ser Ser Lys Ile Ala Thr Asn Gly Gly
      85                      90                      95
Thr Ser Val Glu Gly Leu Gly Phe Ala Ile Pro Ala Asn Asp Ala Ile
      100                      105                      110
Asn Ile Ile Glu Gln Leu Glu Lys Asn Gly Lys Val Thr Arg Pro Ala
      115                      120                      125
Leu Gly Ile Gln Met Val Asn Leu Ser Asn Val Ser Thr Ser Asp Ile
      130                      135                      140
Arg Arg Leu Asn Ile Pro Ser Asn Val Thr Ser Gly Val Ile Val Arg
      145                      150                      155                      160
Ser Val Gln Ser Asn Met Pro Ala Asn Gly His Leu Glu Lys Tyr Asp
      165                      170                      175
Val Ile Thr Lys Val Asp Asp Lys Glu Ile Ala Ser Ser Thr Asp Leu
      180                      185                      190
Gln Ser Ala Leu Tyr Asn His Ser Ile Gly Asp Thr Ile Lys Ile Thr
      195                      200                      205
Tyr Tyr Arg Asn Gly Lys Glu Glu Thr Thr Ser Ile Lys Leu Asn Lys
      210                      215                      220
Ser Ser Gly Asp Leu Glu Ser
      225                      230

```

(2) INFORMATION FOR SEQ ID NO:150:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 40 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:150:

```

Val Gln Arg Ser Met Leu Leu Pro Gly Gly Ile Leu Gly Met Thr Val
  1                      5                      10                      15
Trp Leu Ile Tyr Leu Leu Leu Lys Glu Pro Thr Asn Val Ile Val Ala
      20                      25                      30
Val Asn Gln Ser Leu Lys Arg Ser
      35                      40

```

(2) INFORMATION FOR SEQ ID NO:151:

(i) SEQUENCE CHARACTERISTICS:

239

- (A) LENGTH: 102 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:151:

```

Val Thr Met Glu Leu Asn Thr His Asn Ala Glu Ile Leu Leu Ser Ala
 1           5           10           15
Ala Asn Lys Ser His Tyr Pro Gln Asp Glu Leu Pro Glu Ile Ala Leu
          20           25           30
Ala Gly Arg Ser Asn Val Gly Lys Ser Ser Phe Ile Asn Thr Met Leu
          35           40           45
Asn Arg Lys Asn Leu Ala Arg Thr Ser Gly Lys Pro Gly Lys Thr Gln
          50           55           60
Leu Leu Asn Phe Phe Asn Ile Asp Asp Lys Met Arg Phe Val Asp Val
65           70           75           80
Pro Gly Tyr Gly Tyr Ala Arg Val Ser Lys Lys Glu Arg Glu Lys Trp
          85           90           95
Gly Cys Met Ile Glu Glu
          100

```

(2) INFORMATION FOR SEQ ID NO:152:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 70 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:152:

```

Val Gln Met Tyr Glu Phe Leu Lys Tyr Tyr Glu Ile Pro Val Ile Ile
 1           5           10           15
Val Ala Thr Lys Ala Asp Lys Ile Pro Arg Gly Lys Trp Asn Lys His
          20           25           30
Glu Ser Ala Ile Lys Lys Lys Leu Asn Phe Asp Pro Ser Asp Asp Phe
          35           40           45
Ile Leu Phe Ser Ser Val Ser Lys Ala Gly Met Asp Glu Ala Trp Asp
          240

```

50 55 60
Ala Ile Leu Glu Lys Leu
65 70

(2) INFORMATION FOR SEQ ID NO:153:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 59 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:153:

Val	Phe	Met	Val	Tyr	Asn	Cys	Pro	Lys	Pro	Val	Tyr	Ser	Phe	Leu	Lys
1				5					10					15	
Ser	Ala	Ile	Asn	Leu	Met	Ala	Ala	Ile	Pro	Ser	Ile	Val	Tyr	Gly	Phe
			20					25					30		
Phe	Gly	Leu	Gln	Leu	Leu	Val	Pro	Trp	Ile	Lys	Thr	Phe	Leu	Gly	Asn
		35					40					45			
Gly	Met	Ser	Cys	Pro	Asn	Gln	Leu	Arg	Tyr	Tyr					
	50					55									

(2) INFORMATION FOR SEQ ID NO:154:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 294 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:154:

Val	Ile	Ile	Met	Lys	Phe	Lys	Lys	Met	Leu	Thr	Leu	Ala	Ala	Ile	Gly
1				5					10					15	
Leu	Ser	Gly	Phe	Gly	Leu	Val	Ala	Cys	Gly	Asn	Gln	Ser	Ala	Ala	Ser
			20					25					30		
Lys	Gln	Ser	Ala	Pro	Gly	Thr	Ile	Glu	Val	Ile	Ser	Arg	Glu	Asn	Gly
		35						40					45		

241

```

Ser Gly Thr Arg Gly Ala Phe Thr Glu Ile Thr Gly Ile Leu Lys Lys
 50                               55                               60
Asp Gly Asp Lys Lys Ile Asp Tyr Thr Ala Lys Thr Ala Val Ile Gln
65                               70                               75                               80
Asn Ser Thr Glu Gly Val Leu Ser Ala Val Gln Gly Asn Ala Asn Ala
                               85                               90                               95
Ile Gly Tyr Ile Ser Leu Gly Ser Leu Thr Lys Ser Val Lys Ala Leu
                               100                               105                               110
Glu Ile Asp Gly Val Lys Ala Ser Arg Asp Thr Val Leu Asp Gly Glu
                               115                               120                               125
Tyr Pro Leu Gln Arg Pro Phe Asn Ile Val Trp Ser Ser Asn Leu Ser
                               130                               135                               140
Lys Leu Gly Gln Asp Phe Ile Ser Phe Ile His Ser Lys Gln Gly Gln
145                               150                               155                               160
Gln Val Val Thr Asp Asn Lys Phe Ile Glu Ala Lys Thr Glu Thr Thr
                               165                               170                               175
Glu Tyr Thr Ser Gln His Leu Ser Gly Lys Leu Ser Val Val Gly Ser
                               180                               185                               190
Thr Ser Val Ser Ser Leu Met Glu Lys Leu Ala Glu Ala Tyr Lys Lys
                               195                               200                               205
Glu Asn Pro Glu Val Thr Ile Asp Ile Thr Ser Asn Gly Ser Ser Ala
                               210                               215                               220
Gly Ile Thr Ala Val Lys Glu Lys Thr Ala Asp Ile Gly Met Val Ser
225                               230                               235                               240
Arg Glu Leu Thr Pro Glu Glu Gly Lys Ser Leu Thr His Asp Ala Ile
                               245                               250                               255
Ala Leu Asp Gly Ile Ala Val Val Val Asn Asn Asp Asn Lys Ala Ser
                               260                               265                               270
Gln Val Ser Met Ala Glu Leu Ala Asp Val Phe Ser Gly Lys Leu Thr
                               275                               280                               285
Thr Trp Asp Lys Ile Lys
                               290

```

(2) INFORMATION FOR SEQ ID NO:155:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 158 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:155:

Val	Ser	Ser	Ile	Leu	Gly	Ala	Gly	Pro	Phe	Phe	Gly	Leu	Ala	His	Glu
1				5				10					15		
Ala	Gln	Leu	Lys	Ile	Leu	Glu	Leu	Thr	Ala	Gly	Gln	Val	Ala	Thr	Met
			20					25					30		
Tyr	Glu	Ser	Pro	Val	Gly	Phe	Arg	His	Gly	Pro	Lys	Ser	Leu	Ile	Asn
			35				40						45		
Asp	Asn	Thr	Val	Val	Leu	Val	Phe	Gly	Thr	Thr	Thr	Asp	Tyr	Thr	Arg
			50				55				60				
Lys	Tyr	Asp	Leu	Asp	Leu	Val	Arg	Glu	Val	Ala	Gly	Asp	Gln	Ile	Ala
65					70					75				80	
Arg	Arg	Val	Val	Leu	Leu	Ser	Asp	Gln	Ala	Phe	Gly	Leu	Glu	Asn	Val
				85					90					95	
Lys	Glu	Val	Ala	Leu	Gly	Cys	Gly	Gly	Val	Leu	Asn	Asp	Ile	Tyr	Arg
			100					105					110		
Val	Phe	Pro	Tyr	Ile	Val	Tyr	Ala	Gln	Leu	Phe	Ala	Leu	Leu	Thr	Ser
			115				120					125			
Leu	Lys	Val	Glu	Asn	Lys	Pro	Asp	Thr	Pro	Ser	Pro	Thr	Gly	Thr	Val
			130				135					140			
Asn	Arg	Val	Val	Gln	Gly	Val	Ile	Ile	His	Glu	Tyr	Gln	Lys		
145					150						155				

(2) INFORMATION FOR SEQ ID NO:156:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 271 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:156:

Val	Lys	Pro	Gly	Asp	Phe	Val	Ile	Val	Pro	Phe	Thr	His	Gly	Cys	Gly
1				5				10					15		
Glu	Cys	Asp	Ala	Cys	Leu	Ala	Gly	Phe	Asp	Gly	Ser	Cys	Asp	Asn	His
			20					25					30		
Ile	Gly	Asn	Asn	Leu	Gly	Gly	Asp	Phe	Gln	Ala	Glu	Tyr	Ile	Arg	Phe
			35				40					45			
His	Tyr	Ala	Asn	Trp	Ala	Leu	Val	Lys	Ile	Pro	Gly	Gln	Pro	Ser	Asp
			50				55					60			

243

```

Tyr Thr Glu Gly Met Leu Lys Ser Leu Leu Thr Leu Ala Asp Val Met
65              70              75              80
Pro Thr Gly Tyr His Ala Ala Arg Val Ala Asn Val Gln Lys Gly Asp
              85              90              95
Lys Val Val Val Ile Gly Asp Gly Ala Val Gly Gln Cys Ala Val Ile
              100              105              110
Ala Ala Lys Met Arg Gly Ala Ser Gln Ile Ile Leu Met Ser Arg His
              115              120              125
Glu Asp Arg Gln Lys Met Ala Met Glu Ser Gly Ala Thr Ala Val Val
              130              135              140
Ala Glu Arg Gly Gln Glu Gly Ile Thr Lys Val Arg Glu Ile Leu Gly
145              150              155              160
Gly Gly Ala Asp Ala Ala Leu Glu Cys Val Gly Thr Glu Ala Ala Ile
              165              170              175
Glu Gln Ala Leu Gly Val Leu His Asn Gly Gly Arg Met Gly Phe Val
              180              185              190
Gly Val Pro His Tyr Asn Asn Arg Ala Leu Gly Ser Thr Phe Met Gln
              195              200              205
Asn Ile Ser Val Ala Gly Gly Ala Ala Ser Ala Thr Thr Tyr Asp Lys
              210              215              220
Gln Phe Leu Leu Lys Ala Val Leu Asp Gly Asp Ile Asn Pro Gly Arg
225              230              235              240
Val Phe Thr Ser Ser Tyr Lys Leu Glu Asp Ile Asp Gln Ala Tyr Lys
              245              250              255
Asp Met Asp Glu Arg Lys Thr Ile Lys Ser Met Ile Val Ile Glu
              260              265              270

```

(2) INFORMATION FOR SEQ ID NO:157:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 122 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:157:

```

Val Arg Lys Ser Arg Val Asn Asn Ser Gln Gln Met Leu Gln Ala Leu
1              5              10              15
Glu Glu Gln Asp Leu Thr Lys Ala Glu His Tyr Phe Ala Lys Ala Leu
              20              25              30
              244

```

Glu Asn Asp Ser Ser Asp Leu Leu Tyr Glu Leu Ala Thr Tyr Leu Glu
 35 40 45
 Gly Ile Gly Phe Tyr Pro Gln Ala Lys Glu Ile Tyr Leu Lys Ile Val
 50 55 60
 Glu Glu Phe Pro Glu Val His Leu Asn Leu Ala Ala Met Ala Ser Glu
 65 70 75 80
 Asp Gly Gln Ile Glu Lys Ala Phe Asn Tyr Leu Glu Glu Ile Gln Ala
 85 90 95
 Asp Ser Asp Trp Tyr Val Ser Leu Phe Gly Ser Glu Gly Arg Pro Ile
 100 105 110
 Pro Ala Gly Arg Phe Asp Arg Cys Gly Thr
 115 120

(2) INFORMATION FOR SEQ ID NO:158:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 317 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:158:

Val Thr Gly Met Ser Arg Ser Leu Ala Leu Lys Ala Asp Leu Tyr Gln
 1 5 10 15
 Leu Glu Gly Leu Thr Asp Val Ala Arg Glu Lys Leu Leu Glu Ala Leu
 20 25 30
 Thr Tyr Ser Lys Asp Ser Leu Leu Ile Leu Gly Leu Ala Lys Leu Asp
 35 40 45
 Ser Glu Leu Glu Asn Tyr Gln Ala Ala Ile Gln Ala Tyr Ala Gln Leu
 50 55 60
 Asp Asn Arg Ser Ile Tyr Glu Gln Thr Gly Ile Ser Thr Tyr Gln Arg
 65 70 75 80
 Ile Gly Phe Ala Tyr Ala Gln Leu Gly Lys Phe Glu Thr Ala Thr Glu
 85 90 95
 Phe Leu Glu Lys Ala Leu Glu Leu Glu Tyr Asp Asp Leu Thr Ala Phe
 100 105 110
 Glu Leu Ala Ser Leu Tyr Phe Asp Gln Glu Glu Tyr Gln Lys Ala Thr
 115 120 125
 Leu Tyr Phe Lys Gln Leu Asp Thr Ile Ser Pro Asp Phe Glu Gly Tyr
 130 135 140

245

```

Glu Tyr Gly Tyr Ser Gln Ala Leu His Lys Glu His Gln Val Gln Glu
145                150                155                160
Ala Leu Arg Ile Ala Lys Gln Gly Leu Glu Lys Asn Pro Phe Glu Thr
                165                170                175
Arg Leu Leu Leu Ala Ala Ser Gln Phe Ser Tyr Glu Leu His Asp Ala
                180                185                190
Ser Gly Ala Glu Asn Tyr Leu Leu Thr Ala Lys Glu Asp Ala Glu Asp
                195                200                205
Thr Glu Glu Ile Leu Leu Arg Leu Ala Thr Ile Tyr Leu Glu Gln Glu
                210                215                220
Arg Tyr Glu Asp Ile Leu Asp Leu Gln Ser Glu Glu Pro Glu Asn Leu
225                230                235                240
Leu Thr Lys Trp Met Ile Ala Arg Ser Tyr Gln Glu Met Asp Asp Leu
                245                250                255
Asp Thr Ala Tyr Glu His Tyr Gln Glu Leu Thr Gly Asp Leu Lys Asp
                260                265                270
Asn Pro Glu Phe Leu Glu His Tyr Ile Tyr Leu Leu Arg Glu Leu Gly
                275                280                285
His Phe Glu Glu Ala Lys Val His Ala His Thr Tyr Leu Lys Leu Val
                290                295                300
Pro Asp Asp Val Gln Met Gln Glu Leu Phe Glu Arg Leu
305                310                315

```

(2) INFORMATION FOR SEQ ID NO:159:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 77 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:159:

```

Val Glu Lys Ala Gly Val Val Ile Ala Ile Asn His Asn Glu Ile Pro
 1                5                10                15
Trp Glu Thr Ile Asp Gly Lys Gly Val Lys Val Ile Val Leu Phe Ala
                20                25                30
Val Gly Asp Asp Thr Glu Ala Ala Arg Glu His Leu Lys Thr Leu Ser
                35                40                45
Leu Phe Ala Arg Lys Leu Gly Asn Asp Glu Val Val Ala Lys Leu Val
50                55                60

```

246

Arg Ala Gln Thr Ser Asp Asp Val Ile Ala Ala Phe Cys
 65 70 75

(2) INFORMATION FOR SEQ ID NO:160:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 46 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:160:

Val	Ser	Asp	Phe	His	Asp	Phe	Ser	Asp	Arg	Glu	Val	Arg	Trp	Leu	Ser
1				5					10					15	
Pro	Glu	Glu	Phe	Lys	Asn	Tyr	Pro	Leu	Ala	Lys	Pro	Gln	Gln	Lys	Ile
			20					25					30		
Trp	Gln	Ala	Tyr	Ala	Gln	Ala	Asn	Leu	Asp	Ser	Ser	Gln	Asp		
		35					40					45			

(2) INFORMATION FOR SEQ ID NO:161:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 96 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:161:

Val	Asn	Phe	Glu	Lys	Lys	Ala	Gln	Thr	Gln	Ile	Ala	Gln	Ile	Val	Gln
1				5					10					15	
Asn	Gly	Trp	Asp	Lys	Leu	Pro	Ile	Cys	Met	Ala	Lys	Thr	Gln	Tyr	Ser
			20					25					30		
Phe	Ser	Asp	Asn	Pro	Asn	Ala	Leu	Gly	Ala	Pro	Glu	Asn	Phe	Glu	Ile
		35				40					45				
Thr	Ile	Arg	Glu	Leu	Val	Pro	Lys	Leu	Gly	Ala	Gly	Phe	Ile	Val	Ala
	50					55					60				
Leu	Thr	Gly	Asp	Val	Met	Thr	Met	Pro	Gly	Leu	Pro	Lys	Arg	Pro	Ala
															247

65		70		75		80									
Ala	Leu	Asn	Met	Asp	Val	Glu	Ser	Asp	Gly	Thr	Val	Leu	Gly	Leu	Phe
		85							90					95	

(2) INFORMATION FOR SEQ ID NO:162:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 292 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:162:

Val	Lys	Lys	Arg	Lys	Lys	Leu	Ala	Leu	Ser	Leu	Ile	Ala	Phe	Trp	Leu
1				5					10					15	
Thr	Ala	Cys	Leu	Val	Gly	Cys	Ala	Ser	Trp	Ile	Asp	Arg	Gly	Glu	Ser
			20					25					30		
Ile	Thr	Ala	Val	Gly	Ser	Thr	Ala	Leu	Gln	Pro	Leu	Val	Glu	Val	Ala
		35					40						45		
Ala	Asp	Glu	Phe	Gly	Thr	Ile	His	Val	Gly	Lys	Thr	Val	Asn	Val	Gln
	50					55				60					
Gly	Gly	Ser	Ser	Gly	Thr	Gly	Leu	Ser	Gln	Val	Gln	Ser	Gly	Ala	Val
65				70				75						80	
Asp	Ile	Gly	Asn	Ser	Asp	Val	Phe	Ala	Glu	Glu	Lys	Asp	Gly	Ile	Asp
			85					90					95		
Ala	Ser	Ala	Leu	Val	Asp	His	Lys	Val	Ala	Val	Ala	Gly	Leu	Ala	Leu
			100					105					110		
Ile	Val	Asn	Lys	Glu	Val	Asp	Val	Asp	Asn	Leu	Thr	Thr	Glu	Gln	Leu
		115				120						125			
Arg	Gln	Ile	Phe	Ile	Gly	Glu	Val	Thr	Asn	Trp	Lys	Glu	Val	Gly	Gly
	130				135						140				
Lys	Asp	Leu	Pro	Ile	Ser	Val	Ile	Asn	Arg	Ala	Ala	Gly	Ser	Gly	Ser
145				150					155					160	
Arg	Ala	Thr	Phe	Asp	Thr	Val	Ile	Met	Glu	Gly	Gln	Ser	Ala	Met	Gln
			165					170					175		
Ser	Gln	Glu	Gln	Asp	Ser	Asn	Gly	Ala	Val	Lys	Ser	Ile	Val	Ser	Lys
		180					185						190		
Ser	Pro	Gly	Ala	Ile	Ser	Tyr	Leu	Ser	Leu	Thr	Tyr	Ile	Asp	Asp	Ser
	195					200						205			
Val	Lys	Ser	Met	Lys	Leu	Asn	Gly	Tyr	Asp	Leu	Ser	Pro	Glu	Asn	Ile
															248

```

      210                      215                      220
Ser Ser Asn Asn Trp Pro Leu Trp Ser Tyr Glu His Met Tyr Thr Leu
225                      230                      235                      240
Gly Gln Pro Asn Glu Leu Ala Ala Glu Phe Leu Asn Phe Val Leu Ser
      245                      250                      255
Asp Glu Thr Gln Glu Gly Ile Val Lys Gly Leu Lys Tyr Ile Pro Ile
      260                      265                      270
Lys Glu Met Lys Val Glu Lys Asp Ala Ala Gly Thr Val Thr Val Leu
      275                      280                      285
Glu Gly Arg Gln
      290

```

(2) INFORMATION FOR SEQ ID NO:163:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 71 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:163:

```

Val Gln Pro Thr Gln Ala Glu Gln Pro Ser Thr Pro Lys Glu Ser Ser
  1                      5                      10                      15
Gln Gln Glu Asn Pro Lys Glu Asp Arg Gly Ala Glu Glu Thr Pro Lys
      20                      25                      30
Gln Glu Asp Glu Gln Pro Ala Glu Ala Gln Glu Ile Lys Val Glu Glu
      35                      40                      45
Pro Val Glu Ser Ile Glu Glu Thr Val Ile Gln Pro Val Glu Gln Pro
      50                      55                      60
Lys Val Glu Thr Pro Ala Val
65                      70

```

(2) INFORMATION FOR SEQ ID NO:164:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 465 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:164:

Val	Leu	Leu	Lys	Met	Asp	Gly	Tyr	Arg	Tyr	Val	Gly	Tyr	Leu	Ser	Gly
1				5					10					15	
Asp	Ile	Leu	Lys	Thr	Leu	Gly	Leu	Asp	Thr	Val	Leu	Glu	Glu	Thr	Ser
			20					25					30		
Ala	Lys	Pro	Gly	Glu	Val	Thr	Val	Val	Glu	Val	Glu	Thr	Pro	Gln	Ser
		35					40					45			
Thr	Thr	Asn	Gln	Glu	Gln	Ala	Arg	Thr	Glu	Asn	Gln	Val	Val	Glu	Thr
	50					55				60					
Glu	Glu	Ala	Pro	Lys	Glu	Glu	Ala	Pro	Lys	Thr	Glu	Glu	Ser	Pro	Lys
65				70					75					80	
Glu	Glu	Pro	Lys	Ser	Glu	Val	Lys	Pro	Thr	Asp	Asp	Thr	Leu	Pro	Lys
			85					90					95		
Val	Glu	Glu	Gly	Lys	Glu	Asp	Ser	Ala	Glu	Pro	Ser	Pro	Val	Glu	Glu
		100						105					110		
Val	Gly	Gly	Glu	Val	Glu	Ser	Lys	Pro	Glu	Glu	Lys	Val	Ala	Val	Lys
	115						120				125				
Pro	Glu	Ser	Gln	Pro	Ser	Asp	Lys	Pro	Ala	Glu	Glu	Ser	Lys	Val	Glu
	130					135				140					
Pro	Pro	Val	Glu	Gln	Ala	Lys	Val	Pro	Glu	Gln	Pro	Val	Gln	Pro	Thr
145				150						155				160	
Gln	Ala	Glu	Gln	Pro	Ser	Thr	Pro	Lys	Glu	Ser	Ser	Gln	Gln	Glu	Asn
			165					170					175		
Pro	Lys	Glu	Asp	Arg	Gly	Ala	Glu	Glu	Thr	Pro	Lys	Gln	Glu	Asp	Glu
		180					185					190			
Gln	Pro	Ala	Glu	Ala	Gln	Glu	Ile	Lys	Val	Glu	Glu	Pro	Val	Glu	Ser
	195						200					205			
Lys	Glu	Glu	Thr	Val	Asn	Gln	Pro	Val	Glu	Gln	Pro	Lys	Val	Glu	Thr
	210					215				220					
Pro	Ala	Val	Glu	Lys	Gln	Thr	Glu	Pro	Thr	Glu	Glu	Pro	Lys	Val	Glu
225				230						235				240	
Val	Thr	Ser	Ile	Pro	Gln	Thr	Thr	Arg	Tyr	Glu	Glu	Asp	Leu	Thr	Lys
			245					250					255		
Glu	His	Gly	Thr	Arg	Glu	Val	Val	Lys	Glu	Gly	Lys	Asn	Gly	Ser	Arg
		260					265					270			
Thr	Val	Thr	Thr	Pro	Tyr	Ile	Leu	Asn	Ala	Thr	Asp	Gly	Thr	Thr	Thr
	275					280						285			
Glu	Gly	Thr	Ser	Thr	Thr	Asp	Glu	Ala	Glu	Met	Glu	Lys	Glu	Val	Val
	290					295				300					
Arg	Val	Gly	Thr	Lys	Pro	Lys	Glu	Lys	Leu	Ala	Pro	Val	Leu	Ser	Leu
								250							

```

305              310              315              320
Thr Ser Val Thr Asp Asn Ala Met Leu Arg Ser Ala Arg Leu Thr Tyr
              325              330              335
His Leu Glu Asn Thr Asp Ser Val Asp Val Lys Lys Ile His Ala Glu
              340              345              350
Ile Lys Asn Gly Asp Lys Val Val Lys Thr Ile Asp Leu Ser Lys Glu
              355              360              365
Arg Leu Ser Asp Ala Val Asp Gly Leu Glu Leu Tyr Lys Asp Tyr Lys
              370              375              380
Ile Val Thr Ser Met Thr Tyr Asp Arg Gly Asn Gly Glu Glu Thr Ser
385              390              395              400
Thr Leu Glu Glu Thr Pro Leu Arg Leu Asp Leu Lys Lys Val Glu Leu
              405              410              415
Lys Asn Ile Gly Ser Thr Asn Leu Val Lys Val Asn Glu Asp Gly Thr
              420              425              430
Glu Val Ala Ser Asp Phe Leu Thr Ser Lys Pro Val Asp Val Gln Asn
              435              440              445
Tyr Tyr Leu Lys Val Thr Ser Arg Asp Asn Lys Val Val Ser Pro Pro
450              455              460
Ser
465

```

(2) INFORMATION FOR SEQ ID NO:165:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 152 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:165:

```

Val Gln Leu Tyr Lys Ala Trp Ser Glu Ile Gly Ser Val Val His Thr
 1              5              10              15
His Ser Thr Glu Ala Val Ala Trp Ala Gln Ala Gly Arg Asp Ile Pro
              20              25              30
Phe Tyr Gly Thr Thr His Ala Asp Tyr Phe Tyr Gly Ser Ile Pro Cys
              35              40              45
Ala Arg Ser Leu Thr Lys Asp Glu Val Glu Val Ala Tyr Glu Lys Asp
              50              55              60
Thr Gly Leu Val Ile Val Glu Glu Phe Glu His Arg Gly Leu Asn Pro
              251

```

65		70		75		80									
Val	Glu	Val	Pro	Gly	Ile	Val	Val	Arg	Asn	His	Gly	Pro	Phe	Thr	Trp
				85					90					95	
Gly	Lys	Asn	Pro	Glu	Asn	Ala	Val	Tyr	His	Ser	Val	Val	Leu	Glu	Glu
			100					105					110		
Val	Ser	Lys	Met	Asn	Arg	Phe	Thr	Glu	Gln	Ile	Asn	Pro	Arg	Val	Glu
		115					120					125			
Pro	Ala	Pro	Gln	Tyr	Ile	Leu	Glu	Lys	His	Tyr	Gln	Arg	Lys	His	Gly
	130					135					140				
Pro	Asn	Ala	Tyr	Tyr	Gly	Gln	Lys								
145						150									

(2) INFORMATION FOR SEQ ID NO:166:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 74 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:166:

Val	Val	Lys	Ala	Ile	Gln	Asp	Gly	Lys	Ala	Lys	Leu	Val	Phe	Leu	Ala
1				5				10					15		
His	Asp	Ala	Gly	Pro	Asn	Leu	Thr	Lys	Lys	Ile	Gln	Asp	Lys	Ser	His
		20					25				30				
Tyr	Tyr	Gln	Val	Glu	Ile	Val	Thr	Val	Phe	Ser	Thr	Leu	Glu	Leu	Ile
	35					40					45				
Ile	Ala	Val	Gly	Lys	Ser	Arg	Lys	Val	Leu	Ala	Val	Thr	Asp	Ala	Gly
	50					55					60				
Phe	Thr	Lys	Lys	Met	Arg	Ser	Leu	Met	Glu						
65						70									

(2) INFORMATION FOR SEQ ID NO:167:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 190 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:167:

Val	Ala	Asp	Asp	Asp	Gln	Cys	Ile	Phe	Leu	Cys	His	Asn	His	Arg	Ala
1				5					10					15	
Gln	Glu	Ser	Ile	Glu	Phe	Glu	Lys	Met	Ile	Asp	Gln	Leu	Ser	Lys	Tyr
			20					25					30		
Tyr	Ser	Cys	Arg	Ile	Leu	Thr	Glu	Lys	Asp	Ile	Pro	Ser	Ile	Leu	Ser
		35					40					45			
Leu	Tyr	Glu	Ser	Asn	Pro	Leu	Tyr	Phe	Gln	His	Cys	Pro	Pro	Glu	Pro
	50					55				60					
Asn	Phe	Ala	Thr	Val	Lys	Glu	Asp	Met	Leu	Cys	Leu	Pro	Glu	Gly	Lys
65					70					75					80
Ala	Lys	Ala	Asp	Lys	Phe	Phe	Val	Gly	Phe	Trp	Asn	Gly	Phe	Asp	Leu
				85				90					95		
Val	Ala	Val	Met	Asp	Phe	Val	Tyr	Ala	Tyr	Pro	Asp	Glu	Glu	Thr	Val
			100					105					110		
Phe	Ile	Gly	Leu	Phe	Met	Val	Asp	Gln	Ala	Tyr	Gln	Arg	Lys	Gly	Ile
		115					120				125				
Gly	Ser	His	Ile	Val	Thr	Glu	Ala	Leu	Ala	Tyr	Phe	Ala	Lys	Asn	Phe
	130					135				140					
Arg	Lys	Ala	Arg	Leu	Ala	Tyr	Val	Lys	Gly	Asn	Pro	Gln	Ser	Gln	His
145					150					155					160
Phe	Trp	Glu	Lys	Gln	Gly	Phe	Lys	Ser	Ile	Gly	Cys	Glu	Val	Lys	Gln
				165				170						175	
Glu	Leu	Tyr	Thr	Val	Val	Ile	Val	Glu	Gln	Ser	Leu	Glu	Asp		
			180					185					190		

(2) INFORMATION FOR SEQ ID NO:168:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 215 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:168:

Val Ala Leu Thr Pro Leu Leu Lys Glu Glu Gly Val Ala Asp Ile Pro
1 5 10 15
253

```

Ala Tyr Lys Asp Tyr Tyr Val Pro Met Asn Lys Ala Leu Trp Lys Asp
      20                      25                      30
Leu Glu Leu Lys Lys Ile Ser Lys Gln Glu Leu Val Asn Thr Arg Phe
      35                      40                      45
Ser Arg Leu Phe Ala His Phe Gly Gln Glu Lys Asp Gly Ser Phe Leu
      50                      55                      60
Ala Gln Arg Tyr Gln Phe Tyr Leu Ala Gln Gln Gly Gln Thr Leu Ser
      65                      70                      75                      80
Gly Ala His Asp Leu Leu Asp Ser Leu Ile Glu Arg Asp Tyr Asn Leu
      85                      90                      95
Tyr Ala Ala Thr Asn Gly Ile Thr Ala Ile Gln Thr Gly Arg Leu Ala
      100                     105                     110
Gln Ser Gly Leu Ala Pro Tyr Phe Asn Gln Val Phe Ile Ser Glu Gln
      115                     120                     125
Leu Gln Thr Gln Lys Pro Asp Ala Leu Phe Tyr Glu Lys Ile Gly Gln
      130                     135                     140
Gln Ile Ala Gly Phe Ser Lys Glu Lys Thr Leu Met Ile Gly Asp Ser
      145                     150                     155                     160
Leu Thr Ala Asp Ile Gln Gly Gly Asn Asn Ala Gly Ile Asp Thr Ile
      165                     170                     175
Trp Tyr Asn Pro His His Leu Glu Asn His Thr Gln Ala Gln Pro Thr
      180                     185                     190
Tyr Glu Val Tyr Ser Tyr Gln Asp Leu Leu Asp Cys Leu Asp Lys Asn
      195                     200                     205
Ile Leu Glu Lys Ile Thr Phe
      210                     215

```

(2) INFORMATION FOR SEQ ID NO:169:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 299 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:169:

```

Val Ala Ala Leu Ser Gln Gln Asp Val Pro Lys Ala Leu Ser Cys Leu
  1           5           10           15
Asn Leu Leu Phe Asp Asn Gly Lys Ser Met Thr Arg Phe Val Thr Asp
      20           25           30
                        254

```



```

Leu Leu His Tyr Leu Arg Asp Leu Leu Ile Val Gln Thr Gly Gly Glu
    35                                40                                45
Asn Thr His His Ser Ser Val Phe Val Glu Asn Leu Ala Leu Pro Gln
    50                                55                                60
Lys Asn Leu Phe Glu Met Ile Arg Leu Ala Thr Val Asn Leu Ala Asp
    65                                70                                75                                80
Ile Lys Ser Ser Leu Gln Pro Lys Ile Tyr Ala Glu Met Met Thr Val
    85                                90                                95
Arg Leu Ala Glu Ile Lys Pro Glu Pro Ala Leu Ser Gly Ala Val Glu
    100                                105                                110
Asn Arg Ile Ala Thr Leu Arg Gln Glu Val Ala Arg Leu Lys Gln Glu
    115                                120                                125
Leu Ser Asn Ala Gly Ala Val Pro Lys Gln Val Ala Pro Ala Pro Ser
    130                                135                                140
Arg Pro Ala Thr Gly Lys Thr Val Tyr Arg Val Asp Arg Asn Lys Val
    145                                150                                155                                160
Gln Ser Ile Leu Gln Glu Ala Val Glu Asn Pro Asp Leu Ala Arg Gln
    165                                170                                175
Asn Leu Ile Arg Leu Gln Asn Ala Trp Gly Glu Val Ile Glu Ser Leu
    180                                185                                190
Gly Gly Pro Asp Lys Ala Leu Leu Val Gly Ser Gln Pro Val Ala Ala
    195                                200                                205
Asn Glu His His Ala Ile Leu Ala Phe Glu Ser Asn Phe Asn Ala Gly
    210                                215                                220
Gln Thr Met Lys Arg Asp Asn Leu Asn Thr Met Phe Gly Asn Ile Leu
    225                                230                                235                                240
Ser Gln Ala Ala Gly Phe Ser Pro Glu Ile Leu Ala Ile Ser Met Glu
    245                                250                                255
Glu Trp Lys Glu Val Arg Ala Ala Phe Ser Ala Lys Ala Lys Ser Ser
    260                                265                                270
Gln Thr Glu Lys Glu Val Glu Glu Ser Leu Ile Pro Glu Gly Phe Glu
    275                                280                                285
Phe Leu Ala Asp Lys Val Lys Val Glu Glu Asp
    290                                295

```

(2) INFORMATION FOR SEQ ID NO:170:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 147 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:170:

```

Val Pro Leu Val Ile Leu Met Ile Gly Met Leu Ala Gly Ser Ile Ser
 1           5           10           15
His Gln Val Met His Trp Gly Thr Phe Leu Ala Thr Thr Pro Ile Met
          20           25           30
Leu Val Ala Gly Lys Pro Tyr Ile Gln Ser Ala Trp Ala Ser Phe Lys
          35           40           45
Lys His Asn Ala Asn Met Asp Thr Leu Val Ala Leu Gly Thr Leu Val
          50           55           60
Ala Tyr Phe Tyr Ser Leu Val Ala Leu Phe Ala Gly Leu Pro Val Tyr
65           70           75           80
Phe Glu Ser Ala Gly Phe Ile Leu Phe Phe Val Leu Leu Gly Ala Val
          85           90           95
Phe Glu Glu Lys Met Arg Lys Asn Thr Ser Gln Ala Val Glu Lys Leu
          100          105          110
Leu Asp Leu Gln Ala Lys Thr Ala Glu Val Leu Ser Asp Asp Ser Tyr
          115          120          125
Val Gln Val Pro Leu Glu Gln Val Lys Val Arg Asp Leu Asp Ser Ser
          130          135          140
Ala Ser Arg
145

```

(2) INFORMATION FOR SEQ ID NO:171:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 73 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:171:

```

Val Thr Glu Asn Ala Glu Ala Ala Ala Tyr Phe Thr Asp Gln Val Asp
 1           5           10           15
Ser Ala Ala Val Tyr Val Asn Ala Ser Thr Arg Phe Thr Asp Gly Gly
          20           25           30
Gln Phe Gly Leu Gly Cys Glu Met Gly Ile Ser Thr Gln Lys Leu His
          35           40           45

```

256

Ala Arg Gly Pro Met Gly Leu Lys Glu Leu Thr Ser Tyr Lys Tyr Val
 50 55 60
 Val Ala Gly Asp Gly Gln Ile Arg Glu
 65 70

(2) INFORMATION FOR SEQ ID NO:172:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 94 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:172:

Val Asp Leu Pro Gln Gln Phe His Leu Gly Ser Ile Thr Lys Thr Phe
 1 5 10 15
 Gln Trp Leu Val Asp Ile Asn Asn Leu Val Phe Lys Gly Ser Ile Pro
 20 25 30
 Ile Val Ser Leu Leu Phe Ile Tyr Cys Leu Gly Val Asn Ile Ala Lys
 35 40 45
 Ile Tyr Lys Val Asp Thr Val Ser Ala Gly Leu Val Ser Leu Ala Ser
 50 55 60
 Phe Val Ile Ser Ile Gly Ser Thr Val Thr Lys Ser Phe Pro Leu Ala
 65 70 75 80
 Asn Val Gly Asp Val Lys Leu Asp Gln Ile Leu Thr Trp Asn
 85 90

(2) INFORMATION FOR SEQ ID NO:173:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 330 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:173:

Val Ser Leu Arg Leu Ile Tyr Ser Ile Phe Lys Lys Met Arg Lys Asn
 257

1	5	10	15
Met Lys Ile Ser His Met Lys Lys Asp Glu Leu Phe Glu Gly Phe Tyr			
	20	25	30
Leu Ile Lys Ser Ala Asp Leu Arg Gln Thr Arg Ala Gly Lys Asn Tyr			
	35	40	45
Leu Ala Phe Thr Phe Gln Asp Asp Ser Gly Glu Ile Asp Gly Lys Leu			
	50	55	60
Trp Asp Ala Gln Pro His Asn Ile Glu Ala Phe Thr Ala Gly Lys Val			
65	70	75	80
Val His Met Lys Gly Arg Arg Glu Val Tyr Asn Asn Thr Pro Gln Val			
	85	90	95
Asn Gln Ile Thr Leu Arg Leu Pro Gln Ala Gly Glu Pro Asn Asp Pro			
	100	105	110
Ala Asp Phe Lys Val Lys Ser Pro Val Asp Val Lys Glu Ile Arg Asp			
	115	120	125
Tyr Met Ser Gln Met Ile Phe Lys Ile Glu Asn Pro Val Trp Gln Arg			
	130	135	140
Ile Val Arg Asn Leu Tyr Thr Lys Tyr Asp Lys Glu Phe Tyr Ser Tyr			
145	150	155	160
Pro Ala Ala Lys Thr Asn His His Ala Phe Glu Thr Gly Leu Ala Tyr			
	165	170	175
His Thr Ala Thr Met Val Arg Leu Ala Asp Ala Ile Ser Glu Val Tyr			
	180	185	190
Pro Gln Leu Asn Lys Ser Leu Leu Tyr Ala Gly Ile Met Leu His Asp			
	195	200	205
Leu Ala Lys Val Ile Glu Leu Thr Gly Pro Asp Gln Thr Glu Tyr Thr			
	210	215	220
Val Arg Gly Asn Leu Leu Gly His Ile Ala Leu Ile Asp Ser Glu Ile			
225	230	235	240
Thr Lys Thr Val Met Glu Leu Gly Ile Asp Asp Thr Lys Glu Glu Val			
	245	250	255
Val Leu Leu Arg His Val Ile Leu Lys Ser Thr Thr Ala Cys Leu Asn			
	260	265	270
Met Glu Ile Pro Val Arg Pro Arg Ile Met Glu Ala Glu Ile Ile His			
	275	280	285
Met Ile Asp Asn Leu Asp Ala Ser Met Met Met Met Ser Thr Ala Leu			
	290	295	300
Ala Leu Val Asp Lys Gly Glu Met Thr Asn Lys Ile Phe Ala Met Asp			
305	310	315	320
Asn Arg Ser Phe Tyr Lys Pro Asp Leu Asp			
	325	330	

(2) INFORMATION FOR SEQ ID NO:174:

258

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 137 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:174:

```

Val Trp Lys Lys Lys Lys Val Lys Ala Gly Val Leu Leu Tyr Ala Val
 1             5             10             15
Thr Ile Ala Ala Ile Phe Ser Leu Leu Leu Gln Phe Tyr Leu Asn Arg
          20             25             30
Gln Val Ala His Tyr Gln Asp Tyr Ala Leu Asn Lys Glu Lys Leu Val
          35             40             45
Ala Phe Ala Met Ala Lys Arg Thr Lys Asp Lys Val Glu Gln Glu Ser
          50             55             60
Gly Glu Gln Val Phe Asn Leu Gly Gln Val Ser Tyr Gln Asn Lys Lys
65             70             75             80
Thr Gly Leu Val Thr Arg Val Arg Thr Asp Lys Ser Gln Tyr Glu Phe
          85             90             95
Leu Phe Pro Ser Val Lys Ile Lys Glu Glu Lys Arg Asp Lys Lys Glu
          100            105            110
Glu Val Ala Thr Asp Ser Ser Glu Lys Val Glu Lys Lys Lys Ser Glu
          115            120            125
Glu Lys Pro Glu Lys Lys Glu Asn Ser
          130            135

```

(2) INFORMATION FOR SEQ ID NO:175:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 163 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:175:

```

Val Asp Gly Lys Phe Gly Lys His Val Glu Gln Ile Pro Glu Gly Ala
                          259

```

```

      1             5             10             15
Glu Val Ile Asp Tyr Thr Gly Tyr Ser Ile Ala Pro Gly Leu Val Asp
      20             25             30
Thr His Ile His Gly Tyr Ala Gly Val Asp Val Met Asp Asn Asn Ile
      35             40             45
Glu Gly Thr Leu His Thr Met Ser Glu Gly Leu Leu Ser Thr Gly Val
      50             55             60
Thr Ser Phe Leu Pro Thr Thr Leu Thr Ala Thr Tyr Glu Gln Leu Leu
      65             70             75             80
Ala Val Thr Glu Asn Leu Gly Asn His Tyr Lys Glu Ala Thr Gly Ala
      85             90             95
Lys Ile Arg Gly Ile Tyr Tyr Glu Gly Pro Tyr Phe Thr Glu Thr Phe
      100             105             110
Lys Gly Ala Gln Asn Pro Thr Tyr Met Arg Asp Pro Gly Val Glu Glu
      115             120             125
Phe His Ser Trp Gln Lys Ala Ala Asn Gly Leu Leu Asn Lys Ile Arg
      130             135             140
Leu His Gln Asn Val Met Gly Trp Lys Thr Leu Phe Val Gln Leu Arg
      145             150             155             160
Ala Lys Val

```

(2) INFORMATION FOR SEQ ID NO:176:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 234 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:176:

```

Val Arg Arg Ile Glu Glu Lys Cys Lys Leu Ile Ala Gln Leu Asp Thr
      1             5             10             15
Lys Thr Val Tyr Ser Phe Met Glu Ser Val Ile Ser Ile Glu Lys Tyr
      20             25             30
Val Arg Ala Ala Lys Glu Tyr Gly Tyr Thr His Leu Ala Met Met Asp
      35             40             45
Ile Asp Asn Leu Tyr Gly Ala Phe Asp Phe Leu Glu Ile Thr Lys Lys
      50             55             60
Tyr Gly Ile His Pro Leu Leu Gly Leu Glu Met Thr Val Phe Val Asp
      260

```

```

65              70              75              80
Asp Gln Glu Val Asn Leu Arg Phe Leu Ala Leu Ser Ser Val Gly Tyr
              85              90              95
Gln Gln Leu Met Lys Leu Ser Thr Ala Lys Met Gln Gly Glu Lys Thr
              100              105              110
Trp Ser Val Leu Ser Gln Tyr Leu Glu Asp Ile Ala Val Ile Val Pro
              115              120              125
Tyr Phe Asp Arg Val Glu Ser Leu Glu Leu Gly Cys Asp Tyr Tyr Ile
              130              135              140
Gly Val Tyr Pro Glu Thr Leu Ala Ser Glu Phe His His Pro Ile Leu
145              150              155              160
Pro Leu Tyr Arg Val Asn Ala Phe Glu Ser Arg Asp Arg Glu Val Leu
              165              170              175
Gln Val Leu Thr Ala Ile Lys Glu Asn Leu Pro Leu Arg Glu Val Pro
              180              185              190
Leu Arg Ser Arg Gln Asp Val Phe Ile Ser Ala Ser Ser Leu Glu Lys
              195              200              205
Leu Phe Gln Glu Arg Phe Pro Ala Ser Phe Gly Gln Phe Arg Lys Ala
              210              215              220
Tyr Phe Arg His Phe Leu Arg Leu Gly Tyr
225              230

```

(2) INFORMATION FOR SEQ ID NO:177:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 130 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:177:

```

Val Val Glu Arg Ile Lys Ile Ala Arg Ser Tyr Gly Asp Leu Ser Glu
1              5              10              15
Asn Ser Glu Tyr Glu Ala Ala Lys Asp Glu Gln Ala Phe Val Glu Gly
              20              25              30
Gln Ile Ser Ser Leu Glu Thr Lys Ile Arg Tyr Ala Glu Ile Val Asn
              35              40              45
Ser Asp Ala Val Ala Gln Asp Glu Val Ala Ile Gly Lys Thr Val Thr
              50              55              60
Ile Gln Glu Ile Gly Glu Asp Glu Glu Glu Val Tyr Ile Ile Val Gly
              261

```

```

65              70              75              80
Ser Ala Gly Ala Asp Ala Phe Ala Gly Lys Val Ser Asn Glu Ser Pro
              85              90              95
Ile Gly Gln Ala Leu Ile Gly Lys Lys Thr Gly Asp Thr Ala Thr Ile
              100              105              110
Glu Thr Pro Val Gly Ser Tyr Asp Val Lys Ile Leu Lys Val Glu Lys
              115              120              125
Thr Ala
              130

```

(2) INFORMATION FOR SEQ ID NO:178:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 79 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:178:

```

Val Asp Phe Ile Gly Gly Leu Ser Ala Leu Glu Gln Lys Gly Tyr Gln
 1              5              10              15
Lys Gly Asp Glu Ile Leu Ile Asn Ser Ile Pro Arg Ala Leu Thr Glu
              20              25              30
Thr Asp Lys Val Cys Ser Ser Val Asn Ile Gly Ser Thr Lys Ser Gly
              35              40              45
Ile Asn Met Thr Ala Val Ala Asp Met Gly Arg Ile Tyr Gln Gly Asn
              50              55              60
Gly Lys Ser Phe Arg Tyr Gly Ser Gly Gln Val Gly Cys Ile Arg
65              70              75

```

(2) INFORMATION FOR SEQ ID NO:179:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 130 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:179:

```

Val Val Thr Pro Ala Asn Tyr Asn Thr Pro Ala Gln Ile Val Ile Ala
 1           5           10           15
Gly Glu Val Val Ala Val Asp Arg Ala Val Glu Leu Leu Gln Glu Ala
          20           25           30
Gly Ala Lys Arg Leu Ile Pro Leu Lys Val Ser Gly Pro Phe His Thr
          35           40           45
Ala Leu Leu Glu Pro Ala Ser Gln Lys Leu Ala Glu Thr Leu Ala Gln
 50           55           60
Val Ser Phe Ser Asp Phe Thr Cys Pro Leu Val Gly Asn Thr Glu Ala
65           70           75           80
Ala Val Met Gln Lys Glu Asp Ile Ala Gln Leu Leu Thr Arg Gln Val
          85           90           95
Lys Glu Pro Val Arg Phe Tyr Glu Ser Ile Gly Val Met Gln Glu Ala
          100          105          110
Gly Ile Ser Asn Phe Ile Arg Asp Trp Thr Gly Glu Ser Leu Val Arg
          115          120          125
Phe Cys
130

```

(2) INFORMATION FOR SEQ ID NO:180:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 29 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:180:

```

Val His Pro Thr Gly Pro Thr Pro Ala Thr Glu Thr Val Asp Ser Ile
 1           5           10           15
Pro Gly Phe Glu Ala Pro Gln Glu Ser Val Thr Ile Leu
          20           25

```

(2) INFORMATION FOR SEQ ID NO:181:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 104 amino acids
- (B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:181:

```

Val Pro Thr Val Phe His Lys Ser Ala Gln Val Leu Glu Glu Glu Met
 1           5           10           15
Asn Arg Tyr Gln Pro Asp Phe Val Leu Cys Ile Gly Gln Ala Gly Gly
          20           25           30
Arg Thr Ser Leu Thr Pro Glu Arg Val Ala Ile Asn Gln Asp Asp Ala
          35           40           45
Arg Thr Ser Asp Asn Glu Asp Asn Gln Pro Ile Asp Arg Pro Ile Arg
          50           55           60
Pro Asp Gly Ala Ser Ala Tyr Phe Ser Ser Leu Pro Ile Lys Ala Met
65           70           75           80
Val Gln Ala Ile Lys Lys Lys Asp Tyr Arg Pro Leu Phe Pro Ile Arg
          85           90           95
Gln Gly Leu Leu Ser Ala Ala Ile
          100

```

(2) INFORMATION FOR SEQ ID NO:182:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 128 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:182:

```

Val Leu Gln Val Gly Ser Gln Asp Tyr Val Phe Val Leu Gln Gln Asp
 1           5           10           15
Lys Tyr Thr Ser Val Arg Asp Ile Leu Ser Asp Thr Ile Glu Ala Val
          20           25           30
Glu Tyr Asp Phe Gly Leu Arg Leu Ser Ile Met Leu Gly Gln Val Trp
          35           40           45
Ser Gln Thr Gly His Gln Ala Leu Ser Asp Leu Ile Lys Ala Glu Arg
          50           55           60
Asp Leu Phe Lys Thr Trp Trp Arg Gln Gly His Gln Gly Val His Thr
          264

```

65		70		75		80									
Phe	Ser	Gln	Leu	Tyr	Leu	Trp	Ser	Leu	Gly	Glu	Arg	Leu	Val	Asp	Leu
		85							90					95	
Lys	Pro	Ile	Lys	Glu	Cys	Leu	His	Gln	Met	Ile	Leu	Asp	Gln	Asp	Gln
		100						105					110		
Ile	Gln	Glu	Ile	Ile	Leu	Ser	Leu	Trp	Glu	Asn	Ser	Ala	Val	Leu	Thr
		115					120					125			

(2) INFORMATION FOR SEQ ID NO:183:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 214 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:183:

Val	Arg	Arg	Ser	Asp	Arg	Tyr	Ala	Arg	Glu	Val	Gly	Ala	Asp	Cys	Val
1			5						10					15	
Gly	Glu	Phe	Val	Ser	Ala	Thr	Lys	Thr	Tyr	Pro	Val	Ser	Phe	Ile	Asn
			20					25					30		
Tyr	Lys	Gly	Glu	Glu	Val	Cys	Leu	Asp	Gln	Ala	Pro	Ala	Gly	Ser	Ala
		35				40						45			
Pro	Ala	Ala	Gln	Phe	Met	Asp	Gly	Leu	Ile	Gly	Tyr	Gly	Val	Glu	Gln
	50					55				60					
Leu	Ile	Ser	Thr	Gly	Thr	Cys	Gly	Val	Leu	Ala	Asp	Ile	Glu	Glu	Asn
65				70					75					80	
Ala	Phe	Leu	Val	Pro	Val	Arg	Ala	Leu	Arg	Asp	Glu	Gly	Ala	Ser	Tyr
			85					90					95		
His	Tyr	Val	Ala	Pro	Cys	Arg	Tyr	Met	Glu	Met	Gln	Pro	Glu	Ala	Ile
			100					105					110		
Ala	Ala	Ile	Glu	Glu	Val	Leu	Glu	Asp	Arg	Gly	Ile	Pro	Tyr	Glu	Glu
		115					120					125			
Val	Met	Thr	Trp	Thr	Thr	Asp	Gly	Phe	Tyr	Arg	Glu	Thr	Ala	Glu	Lys
	130					135				140					
Val	Ala	Tyr	Arg	Lys	Glu	Glu	Gly	Cys	Ala	Val	Val	Glu	Met	Glu	Cys
145				150				155						160	
Ser	Ala	Leu	Ala	Ala	Val	Ala	Gln	Leu	Arg	Gly	Val	Leu	Trp	Gly	Glu
			165					170					175		
Leu	Leu	Phe	Thr	Ala	Asn	Ser	Leu	Ala	Asp	Leu	Asp	Gln	Tyr	Asn	Ser
															265

	180		185		190
Arg Asp Trp Gly Ser Glu Pro Phe Asn Lys Ala Leu Lys Leu Ser Leu					
195		200		205	
Ala Ser Val His His Leu					
210					

(2) INFORMATION FOR SEQ ID NO:184:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 136 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:184:

Val	Glu	Asn	Leu	Thr	Asn	Phe	Tyr	Glu	Lys	Tyr	Arg	Val	Tyr	Leu	Thr
1				5					10					15	
Arg	Pro	Arg	Leu	Glu	Leu	Leu	Ala	Val	Val	Thr	Ile	Val	Leu	Xaa	Ala
			20					25					30		
Val	Leu	Val	Phe	Phe	Leu	Asn	Ile	Pro	Gly	Lys	Gly	Val	Leu	Lys	Leu
		35				40						45			
Asp	Asn	Gly	Thr	Ile	Val	Tyr	Asp	Gly	Ser	Leu	Val	Arg	Gly	Lys	Met
	50					55						60			
Asn	Gly	Gln	Gly	Thr	Ile	Thr	Phe	Gln	Asn	Gly	Asp	Gln	Tyr	Thr	Gly
65					70				75					80	
Gly	Phe	Asn	Asn	Gly	Ala	Phe	Asn	Gly	Lys	Gly	Thr	Phe	Gln	Ser	Lys
			85						90					95	
Glu	Gly	Trp	Thr	Tyr	Glu	Gly	Asp	Phe	Val	Asn	Gly	Gln	Ala	Glu	Gly
		100					105					110			
Lys	Gly	Lys	Leu	Thr	Thr	Glu	Gln	Glu	Val	Val	Tyr	Glu	Gly	Thr	Phe
		115				120						125			
Lys	Gln	Gly	Val	Phe	Gln	Gln	Lys								
	130					135									

(2) INFORMATION FOR SEQ ID NO:185:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 53 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single

266

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:185:

Val	Phe	Leu	Lys	Glu	Ser	Cys	Gly	Ser	Gly	Ala	Gln	Ile	Ala	Glu	Thr
1				5				10					15		
Phe	His	Gln	Phe	Gly	Gly	Asp	Tyr	Gly	Phe	Glu	Thr	Thr	Asp	Leu	Asn
			20					25					30		
Phe	Asn	Phe	Ala	Thr	Leu	Arg	Arg	Asn	Arg	Glu	Ala	Tyr	Ile	Asp	Arg
			35					40					45		
Ala	Arg	Ser	Ser	Leu											
			50												

What is claimed is

1. An isolated polynucleotide comprising a polynucleotide sequence selected from the group consisting of:
 - (a) a polynucleotide having at least a 70% identity to a polynucleotide encoding a polypeptide comprising an amino acid sequence of Table 1;
 - (b) a polynucleotide having at least a 70% identity to a polynucleotide encoding a mature polypeptide expressed by the gene contained in the *S. pneumoniae* of the deposited strain that was sequenced to obtain a polynucleotide sequence of Table 1;
 - (c) a polynucleotide encoding a polypeptide comprising an amino acid sequence which is at least 70% identical to an amino acid sequence of Table 1;
 - (d) a polynucleotide which is complementary to the polynucleotide of (a), (b) or (c); and
 - (e) a polynucleotide comprising at least 15 sequential bases of the polynucleotide of (a), (b), (c) or (d).
2. The polynucleotide of Claim 1 wherein the polynucleotide is DNA.
3. The polynucleotide of Claim 1 wherein the polynucleotide is RNA.
4. The polynucleotide of Claim 2 comprising the nucleic acid sequence selected from the group consisting of the nucleic acid sequences set forth in Table 1.
5. The polynucleotide of Claim 2 which encodes a polypeptide comprising an amino acid sequence selected from the group consisting of the amino acid sequences set forth in Table 1.
6. A vector comprising the polynucleotide of Claim 1.
7. A host cell comprising the vector of Claim 6.
8. A process for producing a polypeptide comprising: expressing from the host cell of Claim 7 a polypeptide encoded by said DNA.
9. A process for producing a polypeptide or fragment comprising culturing a host of claim 7 under conditions sufficient for the production of said polypeptide or fragment.
10. A polypeptide comprising an amino acid sequence which is at least 70% identical to an amino acid sequence selected from the group consisting of the amino acid sequences set forth in Table 1.
11. A polypeptide comprising an amino acid sequence selected from the group consisting of the amino acid sequences set forth in Table 1.

12. An antibody against the polypeptide of claim 10.
13. An antagonist or agonist of the activity or expression of the polypeptide of claim 10.
14. A method for the treatment or prevention of disease of an individual comprising: administering to the individual a therapeutically effective amount of the polypeptide of claim 10.
15. A method for the treatment of an individual having need to inhibit a bacterial polypeptide comprising: administering to the individual a therapeutically effective amount of the antagonist of Claim 13.
16. A process for diagnosing a disease related to expression or activity of the polypeptide of claim 10 in an individual comprising:
 - (a) determining a nucleic acid sequence encoding said polypeptide, and/or
 - (b) analyzing for the presence or amount of said polypeptide in a sample derived from the individual.
17. A method for identifying compounds which interact with and inhibit or activate an activity of the polypeptide of claim 10 comprising:

contacting a composition comprising the polypeptide with the compound to be screened under conditions to permit interaction between the compound and the polypeptide to assess the interaction of a compound, such interaction being associated with a second component capable of providing a detectable signal in response to the interaction of the polypeptide with the compound;

and determining whether the compound interacts with and activates or inhibits an activity of the polypeptide by detecting the presence or absence of a signal generated from the interaction of the compound with the polypeptide.
18. A method for inducing an immunological response in a mammal which comprises inoculating the mammal with the polypeptide of claim 10, or a fragment or variant thereof, adequate to produce antibody and/or T cell immune response to protect said animal from disease.
19. A method of inducing immunological response in a mammal which comprises delivering a nucleic acid vector to direct expression of a polypeptide of claim 10, or fragment or a variant thereof, for expressing said polypeptide, or a fragment or a variant thereof *in vivo* in order to induce an immunological response to produce antibody and/ or T cell immune response to protect said animal from disease.

20. A polynucleotide comprising a polynucleotide sequence selected from the group consisting of the the first ten polynucleotides sequences from the top of Table 1.

21. A polypeptide comprising a polypeptide encoded by the polynucleotide of claim 20.

22. The isolated polynucleotide of claim 1 wherein said nucleotide is selected from the group consisting of:

(a) a polynucleotide having at least a 90% identity to a polynucleotide encoding a polypeptide comprising the amino acid sequence of Table 1;

(b) a polynucleotide having at least a 90% identity to a polynucleotide encoding the same mature polypeptide expressed by the gene contained in the *S. pneumoniae* of the deposited strain that was sequenced to obtain a polynucleotide sequence of Table 1;

(c) a polynucleotide encoding a polypeptide comprising an amino acid sequence which is at least 90% identical to the amino acid sequence of Table 1;

(d) a polynucleotide which is complementary to the polynucleotide of (a), (b) or (c); and

(e) a polynucleotide comprising at least 15 sequential bases of the polynucleotide of (a), (b), (c) or (d).

23. The isolated polynucleotide of claim 1 selected from the group consisting of:

(a) a polynucleotide having at least a 95% identity to a polynucleotide encoding a polypeptide comprising the amino acid sequence of Table 1;

(b) a polynucleotide having at least a 95% identity to a polynucleotide encoding the same mature polypeptide expressed by the gene contained in the *S. pneumoniae* of the deposited strain that was sequenced to obtain a polynucleotide sequence of Table 1;

(c) a polynucleotide encoding a polypeptide comprising an amino acid sequence which is at least 95% identical to the amino acid sequence of Table 1;

(d) a polynucleotide which is complementary to the polynucleotide of (a), (b) or (c); and

(e) a polynucleotide comprising at least 15 sequential bases of the polynucleotide of (a), (b), (c) or (d).

24. An isolated polynucleotide comprising a polynucleotide sequence selected from the group consisting of:

(a) a polynucleotide having at least a 50% identity to a polynucleotide encoding a polypeptide comprising the amino acid sequence of Table 1 and obtained from a prokaryotic species other than *S. pneumoniae*;

(b) a polynucleotide encoding a polypeptide comprising an amino acid sequence which is at least 50% identical to the amino acid sequence of Table 1 and obtained from a prokaryotic species other than *S. pneumoniae*; and

(c) a polynucleotide which is complementary to the polynucleotide of (a) or (b).

25. An isolated Streptococcal polypeptide having one of the amino acid sequences given in Table 1.

26. An isolated nucleic acid encoding one of the amino acid sequences of Claim 1 and nucleic acid sequences capable of hybridizing therewith under stringent conditions.

27. Recombinant vectors comprising the nucleic acid sequences of Claim 26 and host cells transformed or transfected therewith.

28. A method of identifying an antimicrobial compound comprising contacting candidate compounds with a polypeptide of Claim 1 and selecting those compounds capable of inhibiting the bioactivity of said polypeptide.

29. Antimicrobial compounds identified by the method of Claim 28.

30. An isolated Streptococcal polypeptide having one of the amino acid sequences given in Table 1.

31. An isolated nucleic acid encoding one of the amino acid sequences of Claim 30 and nucleic acid sequences capable of hybridizing therewith under stringent conditions.

32. Recombinant vectors comprising the nucleic acid sequences of Claim 31 and host cells transformed or transfected therewith.

33. A method of identifying an antimicrobial compound comprising contacting candidate compounds with a polypeptide of Claim 30 and selecting those compounds capable of inhibiting the bioactivity of said polypeptide.

34. Antimicrobial compounds identified by the method of Claim 33.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US97/19226

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) :Please See Extra Sheet.

US CL :Please See Extra Sheet.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 424/130.1, 139.1, 184.1; 435/7.1, 69.3, 320.1, 325; 530/300, 350; 536/23.7

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

APS, MEDLINE, BIOSIS, CA, EMBASE, WPIDS

terms: Streptococcus, pneumoniae, dna, polypeptide, treat, diagnose

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 5,476,929 A (BRILES ET AL) 19 December 1995, see entire document	1-5, 20, 22-24, 26, 31
A	SEVIER et al. Monoclonal Antibodies in Clinical Immunology. Clinical Chemistry. 1981, Vol. 27, No. 11, pages 1797-1806, see entire document	1-34
A	US 4,601,980 A (GOEDDEL ET AL) 22 July 1986, see entire document.	1-34
A	US 5,474,905 A (TAI ET AL) 12 December 1995, see entire document.	1-34

☐ Further documents are listed in the continuation of Box C.
 ☐ See patent family annex.

* Special categories of cited documents:	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A" document defining the general state of the art which is not considered to be of particular relevance	"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E" earlier document published on or after the international filing date	"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&" document member of the same patent family
"O" document referring to an oral disclosure, use, exhibition or other means	
"P" document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search

21 JANUARY 1998

Date of mailing of the international search report

20 FEB 1998

 Name and mailing address of the ISA/US
 Commissioner of Patents and Trademarks
 Box PCT
 Washington, D.C. 20231

Facsimile No. (703) 305-3230

Authorized officer

MARK NAVARRO

Telephone No. (703) 308-0196

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US97/19226

A. CLASSIFICATION OF SUBJECT MATTER:

IPC (6):

A61K 38/00, 39/00, 39/395, 39/40; C07H 21/04; C07K 1/00; C12N 15/00; C12P 21/06; G01N 33/53

A. CLASSIFICATION OF SUBJECT MATTER:

US CL :

424/130.1, 139.1, 184.1; 435/7.1, 69.3, 320.1, 325; 530/300, 350; 536/23.7